BIOCHEMICAL ANALYSIS OF HIV RESTRICTION FACTORS: SINGLE DOMAIN DEOXYCYTIDINE DEAMINASES APOBEC3A AND APOBEC3H

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in the

Department of Microbiology and Immunology

University of Saskatchewan

Saskatoon, SK

By

Robin P. Love

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Microbiology and Immunology

University of Saskatchewan

Health Science Building

107 Wiggins Road

Saskatoon, Saskatchewan, Canada S7N 5E5

ABSTRACT

The APOBEC3 (Apo3) family of proteins are single stranded (ss) DNA cytosine deaminases (C → U). They are grouped into two different structural groups, the single catalytic domain Apo3 enzymes (Apo3A, Apo3C, and Apo3H) and the double catalytic domain Apo3 enzymes (Apo3B, Apo3D, Apo3F, and Apo3G). Apo3G has been implicated in protection from HIV proliferation by becoming encapsidated into budding HIV virions and subsequently mutationally inactivating the synthesized provirus. This largely occurs in the absence of HIV viral infectivity factor (Vif) which mediates the ubiquitination and degradation of Apo3G. Apo3G is a processive enzyme, able to catalyze numerous deaminations in a 5'CCC motif in a single interaction with a substrate. There is a paucity of biochemical data on other Apo3 family members. We performed basic biochemical assays that determined the relative specific activities, processivity, cytosine motif preferences, and binding affinities for DNA, of Apo3A and Apo3H using synthetic DNA substrates in deamination assays. We found Apo3A to be an enzyme with low processivity and Apo3H to be a highly processive enzyme; both of which deaminate a 5'TC motif. Using a reconstituted HIV replication assay we assessed if processivity is needed for efficient restriction of HIV. We were able to demonstrate that each, Apo3G, Apo3A, and Apo3H were able to catalyze deaminations during *in vitro* reverse transcription. The mutation profile of both Apo3A and Apo3H showed that the 5'TC motif preference was less effective compared to Apo3G in triggering missense and nonsense mutations in the HIV protease active site coding sequence.

Nuclear DNA can become deaminated by the related Apo3 family member activation-induced deaminase (AID), when it is present in the nucleus of activated B cells. Apo3A and Apo3H are located in the nucleus but the extent of the damage they cause has only recently been investigated. Here we used an *in vitro* transcription assay to determine the efficiency of Apo3A and Apo3H deamination during transcription and found that, like AID, they are highly capable of causing deaminations during transcription.

Taken together, the results presented here demonstrate that processivity is not necessary for an Apo3 enzyme to catalyze deaminations during HIV reverse transcription and that Apo3A and Apo3H can catalyze deaminations during DNA transcription that could damage host genomic DNA. These results imply a potential cost for maintaining nuclear deaminases.

ACKNOWLEDGEMENTS

I first would like to thank my supervisor Dr. Linda Chelico for accepting me into the lab and creating an environment where I and others could excel academically in science. It was her constant support that allowed me to mature as a scientist and better understand what I wanted to make of my future. I look forward to our continued partnership in the lab pursuing projects together. I would also like to extend my thanks to my committee, Dr. Sidney Hayes, Dr. Peter Howard, and Dr. Kerri Kobryn as well as my external examiner Dr. Yu Luo for their thoughtful and objective view of my work. My committee's input prepared me for the rigours of the publishing world.

I would very much like to thank my lab mates Yuqing Feng, Anjuman Ara, Tayyba Baig, Jonathon Webb, Huixin Xu, and Madison Adolph for lively discussion about all topics, not just science. The dynamics of our lab cannot be beat and that takes effort from everyone and for that I am grateful.

This project would not have been possible without the University of Saskatchewan, Microbiology and Immunology Department, and support staff. Without their help everything grinds to halt in short order.

Thank you to the funding agencies; the National Sciences and Engineering Research Council (NSERC), the Saskatchewan Health Research Foundation (SHRF), and the Canadian Institutes of Health Research (CIHR) for providing for this project.

Finally I must thank and acknowledge my parents, Howard and Joanne Love, as well as my brother Bretton Love. They have seen me through the best and worst of the last two years providing me with the strength to carry on.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

(-)DNA	Negative strand DNA/non coding DNA
(+)DNA	Positive strand DNA/coding DNA
A	Adenine
AID	Activation induced deaminase
AIDS	Acquired immunodeficiency syndrome
Apo3	APOBEC3
Apo3A	APOBEC3A
Apo3G	APOBEC3G
Apo3H	APOBEC3H
APOBEC3	Apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3
ATP	Adenosine triphosphates
bp	Nucleotide base pair
C	Cytosine
CD	Cytosine deaminase domain
CD1	Cytosine deaminase domain 1
CD2	Cytosine deaminase domain 2
cDNA	Complementary DNA
CpG	CpG oligodeoxynucleotides
dA	Deoxyadenosine triphosphate
dC	Deoxycytidine triphosphate
ddA	Dideoxyadenosine triphosphate
ddG	Dideoxyganosine triphosphates
ddH_2O	De-ionized and distilled water
dG	Deoxyguanosine triphosphate
dNTPs	Deoxyribonucleotide triphosphates
dsDNA	Double stranded DNA
dT	Deoxythymidine triphosphate
DTT	Dithiothreitol
Fam	Fluorescein
G	Guanine
Hap	Haplotype
HIV	Human immunodeficiency virus
Ig	Immunoglobulin
IL-4	Interleukin-4
K_d	Apparent dissociation constant
LB	Luria-Bertani broth/medium
LTR	Long terminal repeat
MALS	Multi-angle light scattering
MMR	Mismatch repair
mRNA	Messenger RNA
NC	Nucleocapsid
nt	Nucleotides
P_0	Recombinant virus initial passage
P_1	Recombinant virus amplification passage 1
P_2	Recombinant virus amplification passage 2

P_3	Recombinant virus amplification passage 3
P_4	Recombinant virus amplification passage 4
PAGE	Polyacrylamide gel electrophoresis
PBS	Primer binding site
PNK	Polynucleotide kinase
PPT	Polypurine tract
prot	HIV-1 protease gene
PTC	Premature termination codon
R	Nucleotides A or G
RNaseH	Ribonuclease H domain of HIV-1 RT
rNTPs	Ribonucleotides
RT	Reverse transcriptase
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SOB	Super optimal broth
SOC	Enriched super optimal broth
ssDNA	Single-stranded DNA
T	Thymine
TGF-β	Transforming growth factor β
tRNA	Transfer RNA
U	Uracil
UDG	Uracil-DNA glycosylase (derived from *E. coli*)
Vif	Viral infectivity factor
Vpr	Viral protein R of HIV-2
Vpx	Viral protein X of HIV-2
W	Nucleotides A or T
Z1	Zinc coordinating domain type 1
Z2	Zinc coordinating domain type 2
Z3	Zinc coordinating domain type 3

CHAPTER 1.0

INTRODUCTION

1.1 Overview

Human cells contain a retroelement restriction system consisting of the cytosine deaminases from the APOBEC3 family. These enzymes work by deaminating cytosine to uracil in single stranded (ss)DNA which results in C→T or other types of mutations during replication of the deaminated DNA. This can be a defense against retroelements by causing their mutational inactivation. Retroelement DNA, from retrotransposons or retroviruses such as HIV, can become susceptible to deamination during reverse transcription, replication, and RNA-polymerase mediated transcription, because they involve ssDNA intermediates. Although HIV restriction is the main system studied for the effect of Apo3 enzymes, the action of Apo3 enzymes is largely blocked by HIV Vif which causes their ubiquitination and degradation in the proteasome (Conticello *et al.*, 2003; Marin *et al.*, 2008; Marin *et al.*, 2003; Sheehy *et al.*, 2003; Stopak *et al.*, 2003). Unfortunately human DNA can also be susceptible to deamination during transcription (Love *et al.*, 2012; Peled *et al.*, 2008). Therefore, this form of retroelement defense may come at a cellular cost. This has become more apparent in recent years as members of the Apo3 family have been identified such as Apo3A and Apo3H that localize in both the nucleus and the cytoplasm (Chen *et al.*, 2006; Li and Emerman, 2011; OhAinle *et al.*, 2008). Very recent data has also determined that Apo3G also enters the nucleus after irradiation of the cell (Nowarski *et al.*, 2012). Apo3A is known to interact with foreign DNA, in monocytes, leading to degradation of the DNA through a deamination dependant mechanism (Stenglein *et al.*, 2010) as well as the restriction of retrotransposons (Bogerd *et al.*, 2006a; Bogerd *et al.*, 2006b; Chen *et al.*, 2006), and can cause detrimental off-target deamination of nuclear DNA (Landry *et al.*, 2011) or assist in catabolism of DNA (Suspene *et al.*, 2011a). Apo3H is localized in the cytoplasm in most primates, but has become distributed in the cytoplasm and nucleus in humans and chimpanzees (Li and Emerman, 2011; OhAinle *et al.*, 2008; OhAinle *et al.*, 2006). Apo3H in humans also sustained two independent mutations that have resulted in destabilization of the protein, reducing its half-life (OhAinle *et al.*, 2008). The sudden shift of Apo3H from the cytoplasm to the nucleus may have been sufficient evolutionary pressure to result in the selection of a destabilized Apo3H protein if Apo3H were to have caused numerous deaminations in nuclear DNA.

The reference point for nuclear DNA deaminase research is activation induced cytidine deaminase (AID). AID is closely related in form and catalytic activity to the Apo3 family, but has been identified to play a significantly different role in the cell. AID is a ssDNA cytosine deaminase that functions in the adaptive immune system by catalyzing somatic hypermutation mutation of the Ig gene resulting in affinity maturation of antibody and class switch recombination (*reviewed in* Di Noia and Neuberger, 2007; Peled *et al.*, 2008). While these main functional mutations are beneficial in that they lead to heredity-independent diversity of Ig variable region genes, they can potentially be dangerous for the cell (Hasham *et al.*, 2010). The body has developed mechanisms for controlling aberrant deamination catalyzed by AID (Vuong *et al.*, 2009) and Apo3A (Aynaud *et al.*, 2012), but Apo3H has yet to be investigated in this regard.

1.2 Discovery and HIV-APOBEC host-pathogen interaction

In 2002, research into the requirements for sustainable HIV-1 infection in tissue cultures resulted in the discovery of a new family of deoxycytidine deaminase host restriction factors (Goff, 2003; Ho and Bieniasz, 2008; Sheehy *et al.*, 2003). The focus of study at the time was on HIV Viral Infectivity Factor (Vif) and how in some cell lines, Vif's absence resulted in no change in HIV infection ability while in other cell lines its absence would allow only a single round of viral infection (Pomerantz, 2002). Sheehy *et al*. used Vif deleted HIV-1 (Δ*vif* HIV) to infect both isogenic permissive (CEM-SS) and non-permissive (CEM) cell lines (Sheehy *et al.*, 2002). Common among non-permissive cell types was a protein called CEM15 (now known as Apo3G) after the CEM cell used. Simultaneously, Jarmuz *et al*. was searching for cDNA containing the cytidine deaminase active site sequence from APOBEC1, a related family member that deaminates RNA, and found a family of APOBEC genes concentrated on chromosome 22 (Jarmuz *et al.*, 2002). Jarmuz *et al*. grouped the family of genes as APOBEC1, 2 or 3 based on DNA sequence homology (Jarmuz *et al.*, 2002). Since these discoveries in 2002, an explosion of research has been dedicated to understanding these deaminases and the role they play as host restriction factors.

HIV reverse transcription, in the absence of Vif, is the putative target for Apo3G deamination (Figure 1.1). In a natural infection HIV codes for and carries with it viral infectivity factor Vif, which causes the ubiquitination of Apo3G, thereby preventing its function (Figure

1.1) (Conticello *et al.*, 2003; Marin *et al.*, 2008; Marin *et al.*, 2003; Sheehy *et al.*, 2003; Stopak *et al.*, 2003). In the presence of Vif, the amount of Apo3G is severely reduced in the cytoplasm. Since Apo3G must be encapsidated in virions to carry out its function, the low amount of cellular Apo3G results in less encapsidation and a decreased ability of Apo3G to restrict HIV infection. In laboratory Δ*vif* HIV strains Apo3G has been found to package inside of the HIV virion and render the progeny virus non-infectious (Sheehy *et al.*, 2002). Apo3G cannot provide any protection to an HIV infected cell that is producing virus (Figure 1.1). However, once the infected cell begins production of virus particles, Apo3G is able to bind to the nascent RNA genomes and HIV nucleocapsid protein (NC) and enters into the virions (Bach *et al.*, 2008; Huthoff and Malim, 2007; Wang *et al.*, 2007). Apo3G remains in the virus until the particle infects another cell (Figure 1.1).

As the viral core enters the cytoplasm of the target cell, reverse transcription of the genomic (+)RNA to (-)DNA begins. Apo3G can cause C→U deamination of the single strand intermediate generated (Figure 1.2). These uracil lesions become mutations upon synthesis of the complementary DNA strand by reverse transcriptase (RT) as it synthesizes the (+)DNA to make the dsDNA provirus (Figure 1.2 B). Specifically, the RNA genome is primed by the human $tRNA^{Lys3}$ at the primer binding site (PBS) (Figure 1.2 A) (Coffin, 1997). Reverse transcriptase begins synthesizing (-)DNA while the RNaseH domain of other RT molecules, in the virion, catalyze the degradation of the RNA component of the now RNA-DNA heteroduplex (Figure 1.2 A) (Coffin, 1997). While the RNA genome is degraded, NC chaperones the first strand transfer of the (-)DNA to the 3' R sequence and polymerization of (-)DNA continues (Figure 1.2 A) (Coffin, 1997). The RNaseH activity of RT also exposes sections of single stranded (ss) DNA that Apo3G subsequently deaminates (Figure 1.2 A-B) (Yu *et al.*, 2004). With (-)DNA synthesis complete, (+)DNA synthesis is primed by RNaseH resistant polypurine tracts (PPT) of the RNA-DNA heteroduplex (Figure 1.2) (Coffin, 1997). Synthesis of the (+)DNA ends the opportunity for Apo3G deamination and fixes the uracil lesions as G→A transitions on the (+)DNA strand (Figure 1.2 A-B) (See Appendix 7.1). There is a very low potential for repair of these lesions because the process is occurring in the cytoplasm. Apo3G cannot act as a post-entry barrier and must be packaged in the virion and can only provide protection to subsequently infected cells (Figure 1.1) (Mangeat *et al.*, 2003; Zhang *et al.*, 2003).

Regions of the HIV (-)DNA provirus furthest from the PPT second strand primers remain single stranded for longer. This means that the centre and 5' end of the (-)DNA are single-stranded the longest and most susceptible to Apo3G-mediated deaminations (Suspene *et al.*, 2006). This also coincides with the catalytic orientation specificity of Apo3G (3'→5') where the enzyme scans in both directions but prefers to deaminate cytosines furthest from the 3' end of ssDNA (Chelico *et al.*, 2006). It is thought that the kinetics of reverse transcription along with the processivity characteristic of Apo3G results in a gradient of mutations 3' to the PBS as demonstrated *in vitro* (Feng and Chelico, 2011) and in database studies of clinical HIV strains (Suspene *et al.*, 2006). These induced mutations occur in great numbers and can functionally inactivate HIV (Mangeat *et al.*, 2003; Zhang *et al.*, 2003). Apo3G may cause the gradient of deaminations found in the HIV strains because it is processive and able to catalyze multiple deaminations during a single interaction. An Apo3 enzyme that is not processive has never been tested on an actively reverse transcribing substrate for its ability to cause multiple deaminations. A non-processive Apo3 enzyme will allow for the differentiation between the roles that processivity of the deaminase play versus the kinetics of reverse transcription (i.e. availability of single stranded intermediates) in the accumulation of deaminations.

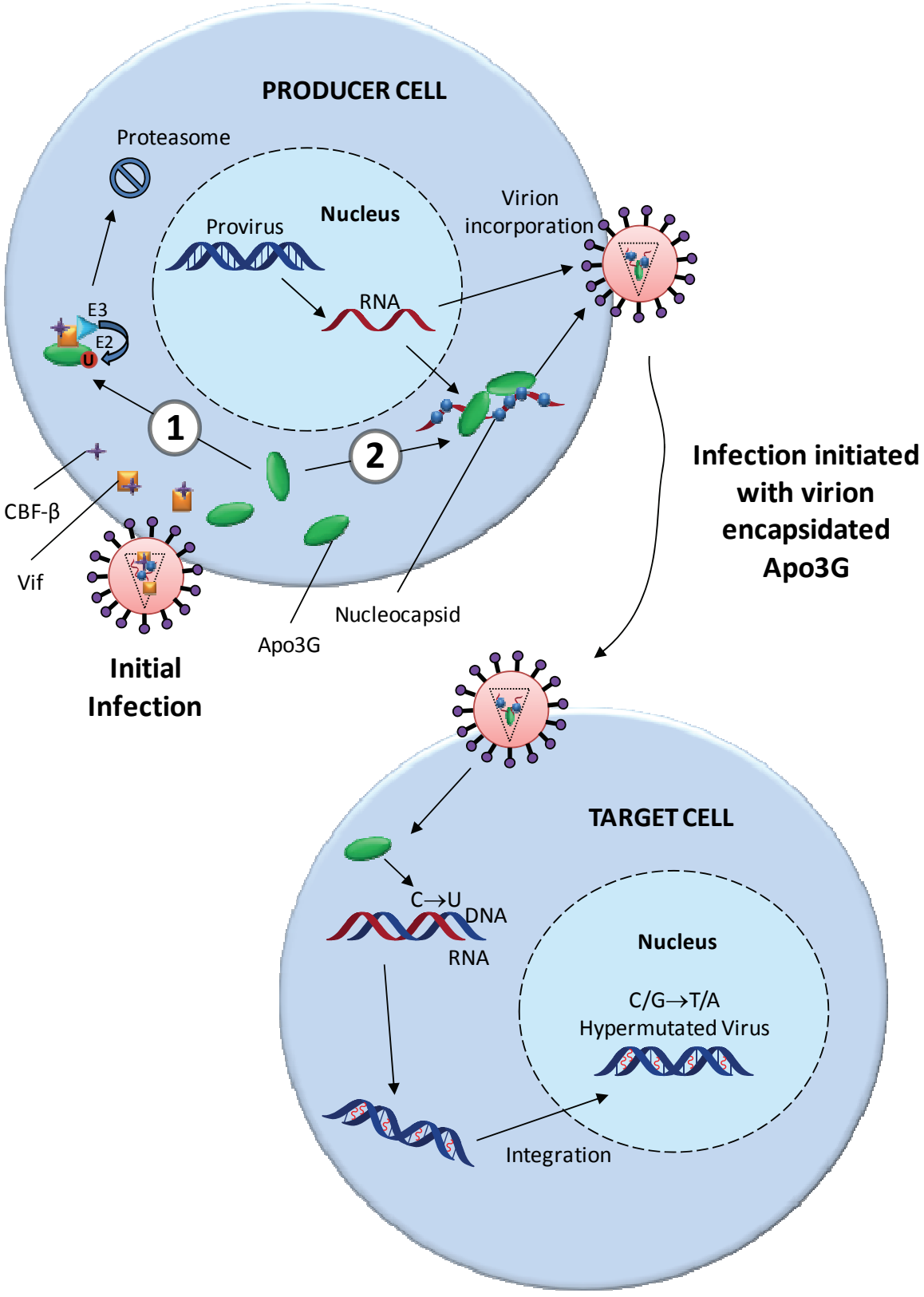
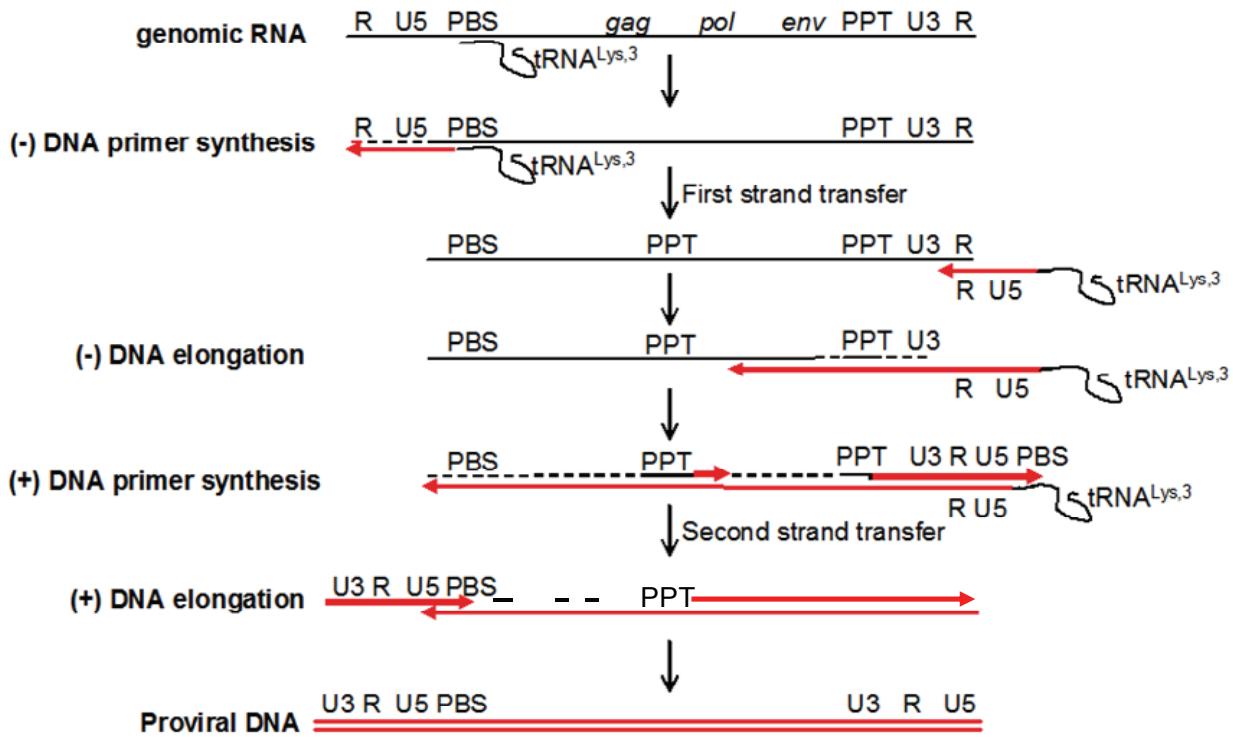

PRODUCER CELL
Proteasome
Nucleus
Provirus
RNA
Virion incorporation
E3
E2
U
1
2
CBF-β
Vif
Apo3G
Nucleocapsid
Initial Infection
Infection initiated with virion encapsidated Apo3G
TARGET CELL
C→U
DNA
RNA
Nucleus
C/G→T/A
Hypermutated Virus
Integration

Figure 1.1 **Apo3G interaction with HIV-1.** (1) Wild type HIV-1 produces viral infectivity factor (Vif) that interacts with the transcription factor CBF-β for stability (Jager *et al.*, 2012; Zhang *et al.*, 2012) and the cellular ubiquitination pathway to cause degradation of Apo3G (Conticello *et al.*, 2003; Marin *et al.*, 2008; Marin *et al.*, 2003; Sheehy *et al.*, 2003; Stopak *et al.*, 2003). As a result, the virus is unaffected by Apo3G inhibition. (2) In a Vif deficient HIV-1 infection, Apo3G is not degraded and remains at a concentration where it binds to HIV genomic RNA and nucleocapsid protein (NC). Apo3G then travels within the virion to the next cell. Upon target cell infection, packaged Apo3G is able to access the initial steps of reverse transcription where it can cause cytosine deaminations (C→U) of (-)DNA. DNA uracil lesions remain unrepaired in the cytoplasm and are fixed as G→A mutations by RT polymerase activity. The provirus can acquire numerous mutations (termed hypermutation) that result in its functional inactivation.

A



B

Genomic RNA	AAAGGGCCUGUCGACCCUUUGGGC	
(-)DNA synthesis	←GGACAGCTGGGAAACCCG	
RNaseH RNA degradation	AA TG A CTT TTTCCCGGACAGCTGGGAAACCCG	RNA degradation occurs nonlinearly
Apo3G catalyzed deamination	TTTUCCGGACAGCTGGGAAAUCCG	
(+)DNA synthesis	AAAAGGCCTGTCGACCCTTTAG→ TTTUCCGGACAGCTGGGAAAUCCG	Seals in G→A mutation

Figure 1.2 **HIV-1 reverse transcription.** A, HIV-1 replication is primed by human $tRNA^{Lys,3}$ binding to the primer binding site (PBS). RNA is depicted in black and DNA is depicted in red. Reverse transcriptase (RT) activity begins (-)DNA synthesis while the RNaseH activity of RT begins degrading the RNA of the RNA-DNA hybrid. Once the RNA is degraded, nucleocapsid protein (NC) chaperones the first strand transfer. Negative strand DNA elongation continues along with RNaseH degradation of the RNA genome with the exception of the two polypurine tracts (PPT) which are RNaseH resistant. Positive strand DNA synthesis is primed by the remaining RNA heteroduplexed PPT regions and carried out by RT. Once (+)DNA elongation is complete NC chaperones the circularization of the proviral DNA (not shown) to complete synthesis. Abbreviations are: R, repeated sequence; U5, unique 5'-region; PBS, primer binding site; PPT, polypurine tract; U3, unique 3'-region. B, Uracil lesions are converted into mutations during HIV reverse transcription. Highlighted nucleotides indicate lesions and subsequent mutations. Panel A adapted from Coffin 1997.

1.3 Apo3 family

The APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3) enzymes are a group of seven enzymes that function in innate immunity (Chiu and Greene, 2008). They are named Apo3A, Apo3B, Apo3C, Apo3DE, Apo3F, Apo3G, and Apo3H and are all located in a gene cluster located on chromosome 22 (Chiu and Greene, 2008; Jarmuz *et al.*, 2002). Of these seven members, three have a single zinc (Zn)-coordinating deaminase domain (Apo3A, Apo3C, and Apo3H) and four have a double zinc-coordinating deaminase domain (Apo3B, Apo3DE, Apo3F, and Apo3G) (Figure 1.3 A) (Bonvin and Greeve, 2007; Chiu and Greene, 2008; Hache *et al.*, 2005; Jarmuz *et al.*, 2002). For the double deaminase domain enzymes, only the C-terminal domain is catalytically active (Jarmuz *et al.*, 2002; Navarro *et al.*, 2005). There are three distinct Z-domain sequences (Z1, Z2, and Z3) (Figure 1.3 B) (LaRue *et al.*, 2009). Apo3 enzymes function both by causing hypermutations through deamination of cytosine to uracil (Mangeat *et al.*, 2003; Zhang *et al.*, 2003), and by binding to RNA of retroelements (Bach *et al.*, 2008; Chiu *et al.*, 2006; Huthoff and Malim, 2007; Wang *et al.*, 2007).

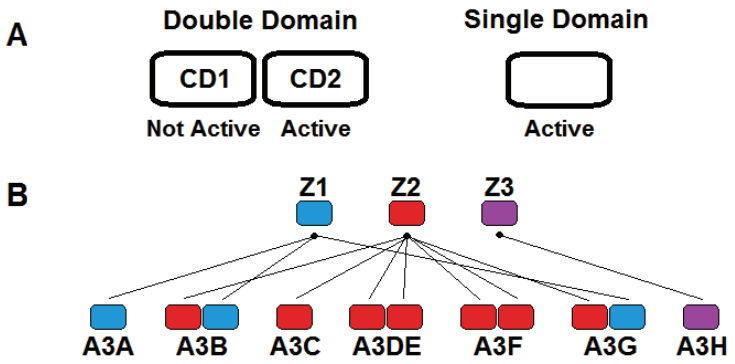


Figure 1.3 **Z deaminase domain expansions.** A, APOBEC3 enzymes are divided into two groups, the single and double Zn-coordinating deaminase domains. In the family members with double deaminase domains, the CD2 domain is catalytically active while the CD1 domain is not. B, This diagram shows the expansion of the Z2 deaminase domain. The Z3 domain was not expanded to other family members. Panel B is adapted from LaRue *et al*. (LaRue *et al.*, 2009).

APOBEC3 enzymes are localized in either the cytoplasm (Apo3F and Apo3G), the nucleus (Apo3B) or both (Apo3A, Apo3C, Apo3DE, Apo3H) (Table 1.1) (Chiu and Greene, 2008). While clinical research on the effectiveness of these enzymes is in its infancy, initial results demonstrate that their up regulation/increased concentration has a positive correlation with better prognoses of patients with HIV (Biasin *et al.*, 2007; Jin *et al.*, 2005; Koning *et al.*, 2009; Land *et al.*, 2008; Pace *et al.*, 2006; Refsland *et al.*, 2010; Ulenga *et al.*, 2008; Vazquez-Perez *et al.*, 2009) and there are many published articles demonstrating their effectiveness *in vitro* against a multitude of retroelements and viruses with ssDNA intermediates including Human Papilloma Virus (Vartanian *et al.*, 2008), Herpes Simplex Virus-1 and Epstein Barr Virus (Suspene *et al.*, 2011b), and parvovirus (Chen *et al.*, 2006). The APOBEC3 enzymes comprise a potentially highly effective defense against HIV, the main model system in which they are studied. Apo3G is the most effective at improving a patient's ability to fight HIV-1 as determined in clinical studies (Biasin *et al.*, 2007; Bishop *et al.*, 2004; Jin *et al.*, 2005; Koning *et al.*, 2009; Land *et al.*, 2008; Pace *et al.*, 2006; Refsland *et al.*, 2010; Ulenga *et al.*, 2008; Vazquez-Perez *et al.*, 2009); however, in most cases it is blocked from acting as a viral restriction factor because of the Vif protein of HIV-1 (Conticello *et al.*, 2003; Marin *et al.*, 2003; Sheehy *et al.*, 2003; Stopak *et al.*, 2003). Vif functions by interacting with Apo3 enzymes (Marin *et al.*, 2008; Sheehy *et al.*, 2002; Sheehy *et al.*, 2003), and causing the ubiquitination and subsequent proteasomal degradation of Apo3G, Apo3F, and Apo3DE most efficiently (Figure 1.1) (Conticello *et al.*, 2003; Marin *et al.*, 2008; Marin *et al.*, 2003; Sheehy *et al.*, 2003; Stopak *et al.*, 2003). In people with an above average Apo3G expression level or who are infected with a Δ*vif* HIV-1, Apo3G can restrict HIV by becoming encapsidated into virions (Albin and Harris, 2010) (Figure 1.2).

Table 1.1 **Physical attributes of various Apo3 family enzymes.**

Enzyme	Molecular Mass	Deamination Motif		Nuclear Localization	Cytoplasmic Localization	Reference
Apo3A	23.0 kDa	5'TTC		~50%	~50%	(Chen *et al.*, 2006)
Apo3H	23.5 kDa	5'CTC	HapI Hap II Hap VII	64% 7.8-40.8% 46.5%	36% 59.2-92.2% 53.5%	(Li and Emerman, 2011; Zhen *et al.*, 2012)
Apo3G	46.4 kDa	5'CCC			>90%	(Chen *et al.*, 2006; Yu *et al.*, 2004)
AID	24.0 kDa	5'WRC	Activated B cell Resting B cell	5-15% >95%	85-95%	(Bransteitter *et al.*, 2004; McBride *et al.*, 2006; Pham *et al.*, 2003)

W denotes A or T; R denotes A or G; underlined C is deaminated; Hap denotes haplotype

Many of the APOBEC3 enzymes have not been assigned a physiological function. The current understanding of these enzymes is that they act primarily in an antagonistic fashion against retroviruses and retroelements (Chiu and Greene, 2008), but other possible cellular roles are only beginning to emerge. For example, recent evidence shows that Apo3G may also act in DNA damage repair by entering the nucleus and interacting with double strand breaks (Nowarski *et al.*, 2012). However, a role in HIV restriction is likely because there are specific viral proteins, HIV-1 Vif and HIV-2 Vpx, which counteract APOBEC3 enzymes. The APOBEC3 proteins could also be contributing to viral evolution by catalyzing an insufficient density of deaminations which result in a virus that is mutated but not functionally inactivated (Jern *et al.*, 2009; Kim *et al.*, 2010; Mulder *et al.*, 2008; Pillai *et al.*, 2008; Sadler *et al.*, 2010).

1.3.1 Apo3G

The first member of the Apo3 family discovered was Apo3G (Sheehy *et al.*, 2002). Apo3G is often used to describe the entire family of Apo3 enzymes, but it is becoming apparent that each of the family members present a unique set of characteristics. Apo3G is highly processive, meaning that it can cause successive deaminations during a single substrate interaction (Chelico *et al.*, 2006; Chelico *et al.*, 2010). The processive nature of Apo3G is due to its double domain structure where the C-terminal CD2 domain is catalytically active for deamination and the N-terminal CD1 domain mediates processivity and polarity (Chelico *et al.*, 2010). Apo3G preferentially catalyzes deaminations from 3' to 5' (Chelico *et al.*, 2006; Chelico *et al.*, 2010). Apo3G scans DNA by facilitated diffusion (Berg *et al.*, 1981; Halford and Marko, 2004) where both jumping and sliding motions occur (Chelico *et al.*, 2006; Chelico *et al.*, 2010; Feng and Chelico, 2011). Apo3G is primarily found in the cytoplasm where it does not come into contact with host DNA (Chiu and Greene, 2008). Apo3G deaminates the motif 5'CC<u>C</u> (underlined C is deaminated), a sequence preference that appears to be unique to Apo3G (Yu *et al.*, 2004). The paradigm set by Apo3G is shown in Figure 1.1 where the Apo3 enzyme must be incorporated into the HIV particle for effective restriction.

Until very recently it was believed that HIV-1 was the sole target for Apo3G activity. A new study has demonstrated that Apo3G is able to enter the nucleus upon DNA damage from radiation (Nowarski *et al.*, 2012). Once in the nucleus, Apo3G localizes to double strand breaks and appears to bind single strand overhangs and bring the ends together (Nowarski *et al.*, 2012).

End joining is dependent on deaminations catalyzed by Apo3G, as demonstrated by the finding that deamination negative mutants are unable to successfully aid in repair (Nowarski *et al.*, 2012). Apo3G was shown to be effective in inducing repair of the double strand breaks through uracil-induced recruitment of DNA repair proteins (Nowarski *et al.*, 2012).

1.3.2 Apo3A

Apo3A is a single domain ssDNA cytosine deaminase that is located in both the nucleus and the cytoplasm (Chen *et al.*, 2006). Apo3A contains the Z1-type zinc coordinating sequence (LaRue *et al.*, 2009) which is in common with the C-terminal catalytic domain of Apo3G (CD1, Z2 binding/processivity; CD2, Z1 catalytic) (Chelico *et al.*, 2010; Feng and Chelico, 2011; Hache *et al.*, 2005; Navarro *et al.*, 2005). It is highly expressed in keratinocytes (Madsen *et al.*, 1999) and in periphery blood monocytes (Chen *et al.*, 2006; Stenglein *et al.*, 2010). Apo3A has a predicted charge of -2.5 at a pH of 7.4, making it unique among Apo3 enzymes which are more close to neutral or positively charged at pH 7.4. No crystal structure has been determined for Apo3A to date, so predictions regarding charge remain sequence based. Apo3A is known to preferentially catalyze deamination of (C→U) in a 5'TTC sequence context from basic assays using purified protein (Chen *et al.*, 2006). Apo3A has not been extensively studied biochemically.

Apo3A is extremely efficient at inhibiting Long Terminal Repeat (LTR) and non-LTR retroelements (Bogerd *et al.*, 2006a; Chen *et al.*, 2006; Muckenfuss *et al.*, 2006). Apo3A has also been shown to be very effective in preventing the replication of Adeno-Associated Virus, a trait not shared with other Apo3 enzymes (Chen *et al.*, 2006). Over expression of Apo3A has been shown to cause hypermutation of Hepatitis B Virus *in vivo* (Abe *et al.*, 2009). Apo3A has been demonstrated to inhibit Human Papilloma Virus types 1 and 16 (Vartanian *et al.*, 2008). Inhibition of HIV-1 virus by Apo3A has recently been demonstrated in dendritic cells (Berger *et al.*, 2011), macrophages (Berger *et al.*, 2011; Koning *et al.*, 2011), and monocytes (Peng *et al.*, 2007) and may contribute in part to the inability of HIV to infect these cells. This activity is unique in the Apo3 family since it appears to occur by Apo3A inhibiting virus upon entry to the target cell. Apo3A does not package into the HIV nucleocapsid which prevents from Apo3A interacting with the ribonucleoprotein complex (Chen *et al.*, 2006; Chiu and Greene, 2008; Goila-Gaur *et al.*, 2007). Agular *et al*. created a fusion protein with Apo3A and HIV Vpr that

was able to enter the viral core (Aguiar *et al.*, 2008). The Apo3A-Vpr fusion protein demonstrated that inhibition of HIV was deamination dependent and required capsid localization. Goila-Gaur *et al*. created protein fusions containing Apo3A fused with the N-terminal half of Apo3G (Goila-Gaur *et al.*, 2007). Similarly, in this experiment they were able to show that the fusion protein was able to inhibit HIV-1 as Apo3A could reach the viral core.

Research by Berger *et al*. has demonstrated that the HIV-2 Vpx protein antagonizes Apo3A in human monocytes (Berger *et al.*, 2010). Targeting by viral factors is a hallmark of a cellular system that evolved to fight viruses and these systems are often highly specific. The interactions observed between Apo3A and HIV-2 indicates that they may have pressured the evolution of each other. It is emerging that Apo3A is a potent antiretroelement and antiretroviral protein where expression of Apo3A is upregulated by interferon-α (Peng *et al.*, 2007; Refsland *et al.*, 2010; Stenglein *et al.*, 2010). Apo3A is also upregulated by CpG DNA (Stenglein *et al.*, 2010).

Along with potent inhibition of endogenous retroelements, Apo3A has been found in high levels in colorectal adenocarcinoma, Burkitt's lymphoma, and chronic myelogenous leukemia cell lines (*See references within* (Chiu and Greene, 2008)). In 2011, Landry *et al*. demonstrated that Apo3A was able to cause genomic instability and subsequent double strand breaks in U2OS cells (an osteosarcoma cell line) (Landry *et al.*, 2011). They found that the damage caused by Apo3A required the catalytic deamination activity of Apo3A and that double strand breaks could only occur in the presence of uracil-N glycosylase (Landry *et al.*, 2011). The DNA damage detected in these experiments was unique to Apo3A as both Apo3C and Apo3G did not demonstrate similar DNA damage (Landry *et al.*, 2011). Landry *et al*. found that DNA replication made nuclear DNA susceptible to Apo3A deamination (Landry *et al.*, 2011). Related family member, AID, has been found to cause aberrant deamination of nuclear DNA during nuclear DNA transcription (Bransteitter *et al.*, 2004; Pham *et al.*, 2003). To date, DNA transcription, which produces transiently single stranded DNA, has not been investigated as a target for Apo3A deamination.

Recently the human tribbles 3 protein (TRIB3) has been identified as being able to interact with Apo3A and modulate its activity(Aynaud *et al.*, 2012). This protein essentially protects nuclear DNA from Apo3A damage. If left unchecked, Apo3A damage begins to

resemble that of uncontrolled AID which causes massive amounts of nuclear DNA damage (Aynaud *et al.*, 2012; Hasham *et al.*, 2010). Together Apo3A and AID demonstrate that nuclear DNA is extremely vulnerable to attack by cellular deaminases.

1.3.3 Apo3H

Apo3H is active against retroviruses such as HIV-1 when overexpressed in tissue culture (Dang *et al.*, 2008; Harari *et al.*, 2009; OhAinle *et al.*, 2008; Tan *et al.*, 2009). Apo3H induces mutations in retroviral DNA by deaminating in a 5'CTC context where the underlined C is preferentially deaminated. Apo3H is a potent inhibitor of MusD and LINE-1 retrotransposons (OhAinle *et al.*, 2008; Tan *et al.*, 2009). Apo3H is an inhibitor of HPV types 1 and 16 (Vartanian *et al.*, 2008). That being said, stable Apo3H is only found in a small fraction of the human population with the vast majority of humans expressing only a form of Apo3H that is quickly degraded (Harari *et al.*, 2009; OhAinle *et al.*, 2008; Wang *et al.*, 2011).

Apo3H was destabilized in human evolution not once, but twice, through two independent mutations (R105G and ΔN15) (OhAinle *et al.*, 2008). These mutations are distributed within the human population resulting in at least seven haplotypes (Hap) (Wang *et al.*, 2011). Both ΔN15 and R105G destabilize the protein, reducing its half life (OhAinle *et al.*, 2008). More recent research has determined that there are subcellular localization differences between the Apo3H haplotypes and that localization appears to be dependent on Arg 105 (Table 1.1). Hap I (15N, 105G, 121K, 178E) is localized in both the nucleus (64%, determined by staining) and cytoplasm (36%) but has a reduced half-life as well as little anti-retroviral activity (Li and Emerman, 2011; Zhen *et al.*, 2012). It appears to enter the nucleus by passive diffusion due to its relatively small size (23.5 kDa) (Li and Emerman, 2011). Hap II (15N, 105R, 121E/D, 178D) remains in the cytoplasm (59.2-92.2%, determined by staining) where it is believed to be actively retained though a mechanism that has not been determined (Li and Emerman, 2011; Zhen *et al.*, 2012). Of particular interest is Hap VII (15N,105R, 121K, 178E) which is used in this M. Sc. research because it contains the ancestral polymorphism 105R, reinstating its stability, and in addition it is partially retained in the nucleus (46.5% nuclear, 53.5% cytoplasmic) where it may interact with nuclear DNA (Li and Emerman, 2011).

The distribution of the haplotypes indicates that ancestral alleles for Apo3H (Hap II) are found in 51-52% of people of African descent, 10-18% of people of European descent, and 3-4% of people of East Asian descent (OhAinle *et al.*, 2008). Despite the decreased stability of Hap I a study in 2011 found that it was able to confer some protection from HIV-1 (Cagliani *et al.*, 2011). Species divergence between humans and chimpanzees predates the acquisition of both the R105G and ΔN15 mutations (OhAinle *et al.*, 2008). It was due to these destabilizing mutations that Apo3H was not initially found along with the other Apo3 family members (Jarmuz *et al.*, 2002; Sheehy *et al.*, 2002), but was discovered later upon sequencing of the human genome (Conticello *et al.*, 2005; Wedekind *et al.*, 2003).

In addition to the two destabilizing mutations (R105G and ΔN15), human Apo3H (all haplotypes) contain a premature termination codon (PTC) that results in the deletion of 29 amino acids at the C-terminal end of the protein that is present in Apo3H of other primates (Dang *et al.*, 2008). OhAinle *et al*. published a report indicating that elimination of the premature termination codon (PTC) drastically changed the subcellular localization of Apo3H Hap I allowing it to localize to the nucleus (OhAinle *et al.*, 2008). They inferred that the PTC removed a putative nuclear export signal and that in its absence Apo3H Hap I had become a nuclear enzyme (OhAinle *et al.*, 2008). Deletion of the PTC of Apo3H allowed for Apo3H retention in the nucleus where it could contact nuclear DNA.

Apo3H is unique among the Apo3 family of cytosine deaminases in terms of its sequence. As discussed in Section 1.3, the Apo3 family contains seven family members (Apo3A-H). Apo3A-G isoforms are each the result of expansion by duplication of the Z1 and Z2 domains while Apo3H is a Z3 type has not been duplicated in all mammals examined (Figure 1.3) (Conticello *et al.*, 2005; LaRue *et al.*, 2009; OhAinle *et al.*, 2006). There are two possible explanations for the absence of expansion of Apo3H. First, it could be that the function of Apo3H predates mammalian evolution and that there was little use for it to remain (OhAinle *et al.*, 2006). This idea was countered by OhAinle *et al*. in 2006 when they found that Apo3H had undergone positive selection, concluding that in recent evolution Apo3H was sustaining mutations at a similar rate to Apo3G (OhAinle *et al.*, 2006; Sawyer *et al.*, 2004), a known HIV-1 inhibitor (Sheehy *et al.*, 2002). A second explanation for the lack of expansion of the Apo3H-Z3

type cytosine deaminase is that it became toxic to the cell due to a subcellular localization change that allowed entry into the nucleus (OhAinle *et al.*, 2008).

Control mechanisms have been identified in the phosphorylation of AID (Vuong *et al.*, 2009) and the binding partner TRIB3 for Apo3A (Aynaud *et al.*, 2012) to restrain the amount of damage done to the host cell but no such mechanism has been found for Apo3H. Though not investigated here, we hypothesise that in lieu of the evolution of a control mechanism in the nucleus for Apo3H, evolution resulted in the inactivation of Apo3H.

Taken together; (i) the strong conservation of destabilizing mutations (R105G and ΔN15) sustained in the vast majority of the population, (ii) the ability of human Apo3H to enter the nucleus where it comes into close contact with nuclear DNA, and finally (iii) the fact that Z3 zinc coordinating domain was not selected for expansion, presents compelling evidence that Apo3H is detrimental to the cell.

1.3.4 Activation induced deaminase

AID is a member of the APOBEC3 type family of ssDNA cytosine deaminases. It has a preference for the 5'WRC motif (W denotes A or T; R denotes A or G; underlined C is deaminated) and deaminates ssDNA processively (Bransteitter *et al.*, 2004; Pham *et al.*, 2003). AID initiates somatic hypermutation during secondary antibody maturation in a transcription dependent manner where U:G mismatches, catalyzed by AID, are repaired by either uracil-N glycosylase or by the mismatch repair pathway (MMR) (*reviewed in* (Di Noia and Neuberger, 2007; Larijani and Martin, 2012; Peled *et al.*, 2008)). Even though AID is only able to deaminate cytosines, diverse amino acid substitutions are observed due to targeted error prone repair involving polη of uracil lesions in antibody diversity regions (McIlwraith *et al.*, 2005). The targeting of the error prone repair remains unknown (Larijani and Martin, 2012). Along with mutations occurring at the amino acid level, AID also causes double strand breaks and deletion of portions of the immunoglobulin cassette that result in antibody class switching (*reviewed in* (Larijani and Martin, 2012; Stavnezer *et al.*, 2008)). Removal of uracil from DNA causes abasic sites that are prone to DNA strand breakage. DNA repair of closely spaced uracil lesions, which are excised to form abasic sites on both strands of the DNA, can results in double strand breaks.

To maintain such a potent deaminase in the nucleus could pose problems for the cell. For this reason AID access to the nucleus is tightly controlled. However, when in the nucleus, AID can cause substantial deaminations in nuclear DNA that are repaired by UNG and downstream repair pathways. Yet, this repair is not always error-free as it has now been demonstrated that most of the B cell chromosome can be edited by AID when it is recruited to stalled PolII RNA polymerase (Liu *et al.*, 2008; Yamane *et al.*, 2011). With the concerted action of UNG, AID can cause massive numbers of double strand breaks which can completely destabilize chromosomes (Hasham *et al.*, 2010). Oncogenes are susceptible to AID deamination because they are highly transcribed due to their role in controlling the cell cycle and many proto-oncogenes have been identified as inadvertent targets for AID (Gordon *et al.*, 2003; Liu *et al.*, 2008; Muschen *et al.*, 2000; Pasqualucci *et al.*, 2001; Robbiani *et al.*, 2009). AID has been found to have a direct link to the transformation of certain cell lines in the absence of the tumour suppressor p53 (Robbiani *et al.*, 2009). The main function of AID to enable somatic hypermutation and class switch recombination is essential to the immune system so this "collateral damage" must be tolerated by the organism and AID cannot be inactivated as is hypothesized for Apo3H which may be functionally redundant with other Apo3 enzymes.

AID can only damage DNA by deamination of cytosine to uracil but the effects can be manifested in two ways, each of which requires DNA repair. Single point mutation repair was investigated by Liu *et al*. who determined that both *Ung* and *Msh2* were needed (Liu *et al.*, 2008). It was determined that the efficiency of Msh2 (MMR type repair) was more essential for preventing damage (Liu *et al.*, 2008). They further determined that AID off target activity was constantly affecting virtually all genes and that mutations were introduced only in specific genes (Ig cassette) due to differences in the repair enzymes recruited to the damaged sites (Liu *et al.*, 2008). In 2010, Hasham *et al*. demonstrated that homologous recombination was required for double strand break repair catalyzed by AID (Hasham *et al.*, 2010). They went on to describe how mutations sustained specifically to XRCC2 homologous recombination factor rendered B cells extremely unstable because AID double strand breaks could not be repaired (Hasham *et al.*, 2010). It has been demonstrated that without high fidelity UNG repair pathways (Liu *et al.*, 2008) for correction of uracil in DNA, AID can cause malignancy in lymphocytes and B cells (Okazaki *et al.*, 2003; Robbiani *et al.*, 2009). It remains to be determined how error prone instead of high fidelity DNA repair is specifically targeted to the Ig genes.

There are several levels of control and regulation pertaining to AID. First, the organism as a whole does not express AID in all tissues, AID is normally only found in B cells (Muramatsu *et al.*, 2000; Muramatsu *et al.*, 1999; Revy *et al.*, 2000). In this way, cell types with large amounts of transcription are free from AID related pressure. Second, there are cellular regulations in place at the protein level. AID is upregulated once a B cell is stimulated for secondary antibody maturation with TGF-β, IL-4, and CD40L (Muramatsu *et al.*, 1999). Tight regulation, mediated by regulatory immune signalling molecules, ensures that risk is reduced and endured only when required. The immune system, in general, exemplifies the evolutionary trade-off between internal sacrifice; i.e. inflammation and cytotoxicity versus defence from external threat. At a cellular level, AID is also found in the cytoplasm where it does not come into contact with cellular DNA when the B cell is at rest (not activated). AID is imported into the nucleus as a result of phosphorylation upon B cell activation (McBride *et al.*, 2006; Vuong *et al.*, 2009). Modification of AID through phosphorylation allows the cell tight control over AID activity. Yet other regulatory molecules are being uncovered. MicroRNA-155 has also been associated with down regulation of AID (Teng *et al.*, 2008).

1.4 Rationale for hypotheses

Apo3A and Apo3H appear to possess HIV restriction ability in cell-based studies (Apo3A, (Berger *et al.*, 2011; Koning *et al.*, 2011); Apo3H, (Dang *et al.*, 2008; Harari *et al.*, 2009; OhAinle *et al.*, 2008; Tan *et al.*, 2009)) although the mechanism by which the enzymes bind ssDNA to scan for cytosines is unknown. However, Apo3A and Apo3H pose a potential problem for cells due to their ability to enter the nucleus (Chen *et al.*, 2006; OhAinle *et al.*, 2008). These deaminases can come into contact with cellular DNA that is transiently single stranded during transcription and replication and is therefore vulnerable to deamination. Cells must have in place systems to handle the potential cellular DNA damage and to avoid incurring a large cellular cost for a mutation-based retroviral restriction system.

1.5 Hypotheses and objectives

This research project biochemically characterized the catalytic activity of single domain Apo3 enzymes Apo3A and Apo3H, on synthetic ssDNA, nascent reverse transcribed DNA and transiently formed ssDNA during gene transcription. This project aims to broaden the

understanding of the Apo3 family and question many of the generalities formed based on research focused on Apo3G.

The hypotheses addressed in this project are:

(1) Processivity is required for efficient restriction of HIV and enzymes not possessing this attribute will catalyze markedly fewer deaminations.
(2) Apo3 enzymes that can localize to the nucleus may cause deamination of nuclear DNA during times when it is transiently single stranded during transcription and replication. This means that there is a cost versus benefit ratio that would impinge selective pressures on the evolution of a mutation-based viral restriction strategy.

CHAPTER 2.0

MATERIALS AND METHODS

2.1 Protein production

Protein production was performed in eukaryotic *Sf9* insect cells because previously analyzed activity, demonstrated that recombinant Apo3G (unpublished data) and AID (Pham *et al.*, 2008) produced from *Escherichia coli* is 100-fold less active than *Sf9* cell expressed Apo3G or AID, likely due to the inability of *E. coli* to correctly phosphorylate the proteins (Demorest *et al.*, 2011).

2.1.1 Production of recombinant baculovirus

Recombinant proteins were produced as GST fusion proteins using a baculovirus expression system. Genes of interest were cloned into the transfer vector pAcG2T (BD Biosciences) by Dr. Chelico using XhoI and EcoRI sites. To create a baculovirus for *Sf9* cell (Life Technologies) infection, the transfer vector, pAcG2T, and 5 μL baculovirus DNA were mixed and incubated for 5 min at room temperature. Then, 100 μL of Sf900 II media (Life Technologies) was added to the DNA mixture and incubated for another 5 min. To facilitate uptake of the DNA mixture into the cell, we used Cellfectin II Reagent (Life Technologies). Cellfectin was prepared by incubating it in 100 μL Sf900 II media for 15 min at room temperature. The Cellfectin and DNA mixture were combined and gently mixed and then incubated for 35 min. A 6-well adherent growth plate was prepared with 1 x 10^6 cells/well in 2 of the wells, one for production of virus and one as a growth control. 2 mL of Sf900 II media was added to cell containing wells. The Cellfectin/DNA mixture was then added carefully drop by drop to the production well. The remaining 4 wells of the 6-well plate were filled with sterile ddH_2O to maintain humidity. The cells were incubated for 4-5 days at 27°C. The supernatant was transferred to a 1.5 mL tube and centrifuged for 15 min at 100 x *g*. Once complete, the supernatant was harvested and saved as the initial passage (P_0) of the virus. To passage the virus, we prepared 3 T75 cm^2 flasks containing 8.4 x 10^6 cells in 10 mL of Sf900 II media for each. Cells were allowed to adhere for 15 min, then the media was replaced with 15 mL fresh Sf900 II media. To the first flask, 100 μL of P_0 was added, to the second 500 μL of P_0 was used, and to the last flask no virus passage was added so that it could serve as a growth control. Infection was monitored for 3 days at 27°C by visual inspection for infection of cells (large swollen cells, detachment, and lack of growth). Virus was passaged (P_1) using the lowest volume that demonstrated infection (either 100 μL or 500 μL). Virus was passaged 2 times more

using 100 μL of supernatant (P_2, P_3). The next step was to amplify the virus for the bulk inoculation. To make the P_4, we prepared a 50 mL suspension flask of cells at 2 x 10^6 cells/mL and inoculated it with 100 μL of P_3. The flask was incubated for 48 hours at 27°C. The solution was then filter sterilized and stored at 4°C. This produces a virus stock of approximately 1 x 10^8 pfu/mL for a 1500 bp gene or 1 x 10^{10} pfu/mL for a 750 bp gene as determined by plaque assays previously done in our lab.

2.1.2 Protein Expression

The optimal amount of virus to add to a culture for protein is empirically determined by conducting small scale infections and purifications. *Sf9* cells are infected with virus when they reach a cell density of 2 x 10^6 cells/mL in suspension culture. Cells are grown on a rotary shaker at 27°C at 122 rpm. It was determined that an MOI of 1 was optimal for Apo3A and Apo3H Hap VII (referred to as Apo3H for simplicity) for 72 h. It has been previously determined that an MOI of 1 for 72 h is optimal for Apo3G expression (Chelico *et al.*, 2008). Cells were pelleted at 100 x *g* for 15 min and washed with phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) were frozen at -80°C for purification at a later time.

2.1.3 Protein purification

A cell pellet from 1 L of infected *Sf9* culture was defrosted on ice for 45 min and resuspended in 20 mL of lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 10 mM NaF, 10 mM $NaH_2PO_4 \cdot H_2O$, 10 mM $Na_4O_7P_2 \cdot 10H_2O$, 10% (v/v) glycerol, 10 mM DTT, 1 protease inhibitor tablet (Roche)) with fresh DTT added to 10 mM along with 1 protease inhibitor tablet (Roche), and 40 μg/mL of RNase A (Qiagen). The mixture was then gently rocked at 4°C for 45 min. An additional 40 μg/mL of RNase A was added before the lysate was sonicated for 1.5 min at 50% output on Setting 5 (Branson Sonifier 450). Sonication ensured complete cell lysis and sheared nucleic acids, especially RNA which is known to purify with Apo3G and AID enzymes (Bransteitter *et al.*, 2003). Cellular debris was pelleted by centrifugation at 4°C for 1 hour at 20 000 x *g*. In the mean time, 1.5 mL of Glutathione Sepharose 4B beads (GE Healthcare) were washed twice with 40 mL PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) by spinning for 1 min at 500 x *g*. Lysate was then removed from the pellet, 20 μg/mL of RNase A added, and incubated with the washed beads for

2 hours at 4°C on a gyro rocker. Beads were then spun for 15 min at 500 x *g* and the lysate was discarded. The beads were then resuspended in PBS Suspension buffer (10 mL 1% Triton X-100, 1X PBS, 500 mM NaCl) and transferred to a 20 mL disposable filter column (Bio-Rad). The filter column was then washed with consecutive 10 mL volumes of modified chilled PBS buffers as follows: PBS Suspension buffer; 1 M NaCl in 1X PBS; and 500 mM NaCl in 1X PBS. The final wash was a 10 mL of Digestion buffer (50 mM Tris pH 7.4, 250 mM NaCl, 1 mM DTT, 10% glycerol). The GST fusion proteins were then digested with Thrombin (25 units/mL) for 16 h at 4°C (Apo3A and Apo3H) or 21°C (Apo3G). The overnight digestion was collected and additional digestion buffer was used to elute the protein. A Bradford assay (BioRad) was used to determine when no additional protein was eluting from the beads. The eluted fractions were concentrated using centrifugal concentrators (Millipore) by using consecutive 10 min centrifugation at 2000 x *g* until an appropriate volume was reached. Sample was mixed between centrifugations to minimize precipitation. Protein purity was confirmed with SDS-PAGE and determined to be ≥95% (Figure 2.1).

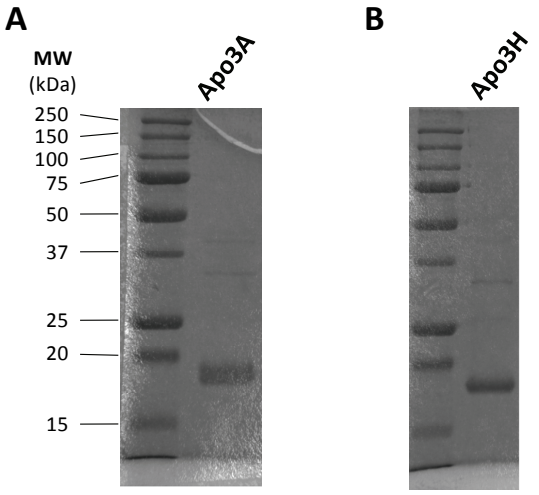


Figure 2.1 **Apo3A and Apo3H enzyme purity verification.** Apo3A and Apo3H were resolved on a 12% SDS-PAGE gel and stained with coomassie. A, Apo3A (23.0 kDa) resolves as a doublet of two isoforms ~19 kDa in size consistent with Chen *et al*. 2006 and Bogerd *et al*. 2006a. B, Apo3H (23.5 kDa) resolves as a ~18 kDa protein. This gel demonstrates that the enzymes are >95% pure.

2.2 Site-directed mutagenesis

The sequence available for Apo3H (NIH accession bc069023) was obtained as a clone (OpenBio Systems) and contained the R105G mutation that results in a reduced half-life (OhAinle *et al.*, 2008). The mutation was reversed by using site directed mutagenesis from the 2-step protocol of Wang and Malcolm (based on the QuikChange Site-Directed Mutagenesis Kit (Stratagene)) (Wang and Malcolm, 1999). This created an Apo3H Hap VII clone (Wang *et al.*, 2011). Primers used were: Apo3H G105R forward primer 5' GCT CAC GAC CAT CTG AAC CTG CGC ATC TTC GCC TCC CGC CTG and Apo3H G105R reverse primer 5' CAG GCG GGA GGC GAA GAT GCG CAG GTT CAG ATG GTC GTG AGC.

Two initial PCR reactions were set up with the template. The reactions were primed in the forward direction or the reverse direction as follows: dNTPs, 10 mM; Apo3H G105R SDM forward primer or Apo3H G105R SDM reverse primer, 0.5 μM; pAcG2T-A3H template 50 ng; 5X Phusion HF buffer (Stratagene), 1X; Phusion DNA polymerase (NEB), 1U; ddH_2O was added to a final volume of 50 μL. The first PCR cycle was set up as follows: 98°C for 60 sec to denature and 98°C for 10 sec, 68°C for 10 sec, 72°C for 2.5 min cycled 2X. 25 μL of the initial PCR reaction with forward primer was mixed with 25 μL of the initial PCR reaction with the reverse primer. An additional 1U of Phusion DNA polymerase was added and the mixture was then subjected to another PCR reaction: 98°C for 60 sec to denature and 98°C for 10 sec, 68°C for 10 sec, 72°C for 2.5 min cycled 29X. The parental template DNA was digested with 10U of *Dpn* I (NEB) for 1 hour at 37°C. DH5α cells were transformed by the heat shock method and select transformed colonies were grown in order for the contained plasmid to be purified. Sequencing was done at the National Research Council (Saskatoon, Saskatchewan).

2.3 Deamination assay

DNA substrates were designed based on those used in similar assays performed by Chelico *et al.* 2006 and were synthesized by TriLink Biotechnologies with an internal fluorescein label as indicated in Table 2.1. Apo3A, Apo3H, and Apo3G were reacted at 20-50 nM with 100 nM or 500 nM fluorescein (Fam)-labeled ssDNA in Reverse Transcriptase Buffer (50 mM Tris, pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, 1 mM DTT). Reactions were incubated for 2.5-30 min at 37°C. Reactions were stopped between 2-30 minutes by mixing 10 μL reaction with 30 μL

phenol chloroform isoamyl alcohol (25:24:1) and 20 μL ddH_2O and vortexing for 30 sec. DNA was extracted and excess phenol was cleared with a chloroform extraction using Phase Lock Tubes as per manufacturer's instruction (5-Prime). DNA samples were then incubated with uracil-DNA glycosylase (UDG, New England Biolabs) for 60 min at 37°C to allow for uracil lesions to become abasic sites. Samples where then heated to 85°C for 10 min with 0.2 mM NaOH to induce breakage at abasic sites for detection of deamination (Chelico *et al.*, 2006). 10μL of DNA samples were run on an 8 cm acrylamide denaturing gel (10%, 16%, or 20% Acrylamide:Bis-Acrylamide 19:1, 8 M urea, 1X TBE Buffer (90 mM Tris, 90 mM Boris acid, 2 mM EDTA pH 8.0)) in 1X TBE Buffer at 18.8 V/cm until the Xylene Cyanol Blue dye reached three quarters of the length of the gel. Gel band intensities were determined with a Typhoon Trio (GE Healthcare) multipurpose scanner or Pharos FX Scanner (BioRad) and integrated gel band intensities were quantified using ImageQuant software (GE Healthcare). The specific activity was determined under single-hit conditions (Creighton *et al.*, 1995) by calculating the pmoles substrate used per minute for a microgram of enzyme.

2.3.1 Processivity analysis

The processivity factor was calculated as previously described (See Appendix 7.2) (Chelico *et al.*, 2006; Chelico *et al.*, 2009). In brief, the value is determined by comparing the deaminations which occurred at two sites on an ssDNA substrate (5'C & 3'C) to the calculated expected value of deaminations which would occur at both motifs if the events were independent (not processive). The expected value was calculated using the probability rule that the frequency of two independent events occurring is the product of their frequencies. The two events considered are the total deaminations occurring at the 5'-proximal C and at the 3'-proximal C. Reactions were performed under single hit conditions to ensure that a single Apo3 was interacting with a given ssDNA substrate.

Table 2.1 **Substrates used for deamination and rotational anisotropy assays.**

Name	Sequence
L3 TTC DNA (deamination assay)	AAA GAG AAA GTG ATA **TTC** A [Fam-dT] A**TT C**AT AGA GTA AAG TTA GTA AGA TGT GTA AGT ATG TTA A
L12 TTC DNA (deamination assay)	AAA GAG TTA GTG AGA **TTC** AAA ATT [Fam-dT] AGA GA**T TC**A AAT GTT AGA TAT GTT AAT GTG TGT GAT GAT GTT GA
L28 TTC DNA (deamination and rotational anisotropy	AAA GAG AAA GTG ATA **TTC** AAA GAG TAA AGT [Fam-dT] AGA TAG AGA GTG ATA **TTC** AAA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
L28 TTC DNA Complement (deamination assay)	TTA ACA TAC TTA CAC ATC TTA CTA ACT TTA CTC TTT GAA TAT CAC TCT CTA TCT AAC TTT ACT CTT TGA ATA TCA CTT TCT CTT T
L28 TTC RNA Complement (deamination assay)	UUA ACA UAC UUA CAC AUC UUA CUA ACU UUA CUC UUU GAA UAU CAC UCU CUA UCU AAC UUU ACU CUU UGA AUA UCA CUU UCU CUU U
L61 TTC DNA (deamination assay)	GAA TAT AGT TTT TAG **TTC** AAA GTA AGT GAA GAT AAT [Fam-dT] TAG AGA GTT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA **TTC** AAA GAG TAA AGT TGT TAA TGT GTG TAG ATA TGT TAA
L28 CTC DNA (deamination assay and rotational anisotropy)	AAA GTG AAA GTG ATA **CTC** AAA TTT AAA AGT [Fam-dT] AGA TAG AAG GTG ATA **CTC** AAA TAT GAA AGT TAG TAA GAT GTG TAA GTA TGT TAA
L28 CTC Reversed Orientation (deamination assay)	AAA TAT GAA AGT TAG TAA GAT GTG TAG ATA AGT TAA ATA **CTC** AAA TTT AAA AGT [Fam-dT] AGA TAG AAG GTG ATA **CTC** AAA GTG AAA GTG
L61 CTC DNA (deamination assay)	GAA TAT AGT TTT TAG **CTC** AAA GTA AGT GAA GAT AAT [Fam-dT] TAG AGA GTT GTA ATG TGA TAT ATG TGT ATG AAA GAT ATA AGA **CTC** AAA GTG AAA AGT TGT TAA TGT GTG TAG ATA TGT TAA
L61 DNA Block (deamination assay)	CTT TCA TAC ACA TAT ATC AC
L61 RNA Block (deamination assay)	CUU UCA UAC ACA UAU AUC AC
L28 CCC DNA (deamination assay)	AAA GAG AAA GTG ATA **CCC** AAA GAG TAA AGT [Fam-dT]AGA TAG AGA GTG ATA **CCC** AAA GAG TAA AGT TAG TAA GAT GTG TAA GTA TGT TAA

Abbreviation: Fam is fluorescein.

2.4 Steady state rotational anisotropy

Steady state fluorescence depolarization (rotational anisotropy) was measured for Apo3A, Apo3H, and Apo3G binding to Fam-labeled ssDNA specific for Apo3A (5'TTC), Apo3H (5'CTC), and Apo3G (5'CCC) (Table 2.1). Reactions were calculated to a volume of 80 µL and contained Fam-labeled ssDNA (15 nM) in Reverse Transcriptase Buffer with a titration of Apo3A (0-25 µM), Apo3H (0-0.45 µM) or Apo3G (0-0.70 µM) to achieve saturation of the substrate. Measurements were made at 21°C using a QuantaMaster QM-4 spectrofluorometer (Photon Technology International) with a dual emission channel. Samples were excited with vertically polarized light at 494 nM (7 nm band pass) and vertical and horizontal emissions were measured at 520 nM (7 nm band pass). Samples were excited for 10 sec at 1 emission measurement recorded per 1 sec and averaged for each titration concentration.

2.5 Multi-angle light scattering assay

Apo3A (300 µg) was subjected to size exclusion chromatography using a Superdex 200HR10/300 column (GE Healthcare) connected to an Agilent 1200 HPLC. A solution containing 20 mM HEPES pH 7.3 and 150 mM NaCl was used as the elution buffer. Chromatography and detection was performed as previously reported (Feng and Chelico, 2011) at the Keck Foundation Biotechnology Resource Laboratory at Yale University (Folta-Stogniew and Williams, 1999). Data analysis to determine molecular masses was performed with ASTRA software (Wyatt, 1993).

2.6 HIV replication assay

2.6.1 Reverse transcription reaction

The model *in vitro* HIV replication assay and subsequent detection of Apo3 catalyzed deamination was performed as described previously (Figure 2.2) (Feng and Chelico, 2011). A synthetic (+) RNA genome was created using *in vitro* RNA transcription from the vector pET-Blue-PPT-Prot (Feng and Chelico, 2011) that contained a polypurine tract (PPT), 120 nt of the catalytic domain of the HIV-1 *protease* (*prot*), and the *lacZα* complementation fragment. The PPT served as an internal second strand ((+) DNA) primer for the reaction (Coffin, 1997). The *prot* was used to identify specific amino acid changes that could be induced by Apo3 enzyme-

mediated deamination, and the *lacZα* was used as a reporter gene for mutations by blue/white screening. The HIV-1 clone from which the protease gene was amplified was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: p93TH253.3 from Dr. Feng Gao and Dr. Beatrice Hahn (Gao *et al.*, 1996).

To begin the reaction, the RNA genome (50 nM) was heat annealed to a 24-nt DNA primer (Feng and Chelico, 2011) (Table 2.2) and incubated with NC (3.0 μM), RT (2.4 μM), and dNTPs (500 μM) in Reverse Transcriptase Buffer in the presence or absence of 200 nM of Apo3A or Apo3G (Figure 2.2). Each of the components of the reverse transcription reactions were at estimated physiological ratios to the HIV genome as found in the HIV-1 literature (Briggs *et al.*, 2004; Coffin, 1997; Zhu *et al.*, 2003) or Apo3G literature (Xu *et al.*, 2007). The reaction mixture was incubated at 37°C for 3 hours. Reverse transcription was stopped and the NC protein filament denatured by the Stopping Solution (0.5% SDS and 2.5 mM EDTA) and proteinase K (0.5 mg/mL) (New England Biolabs). The halted reaction was then incubated at room temperature for 30 min. After the incubation, the DNA was extracted by phenol:chloroform:isoamyl extraction. Excess dNTPs and phenol were removed by using the Bio-Spin 6 Kit Columns (BioRad).

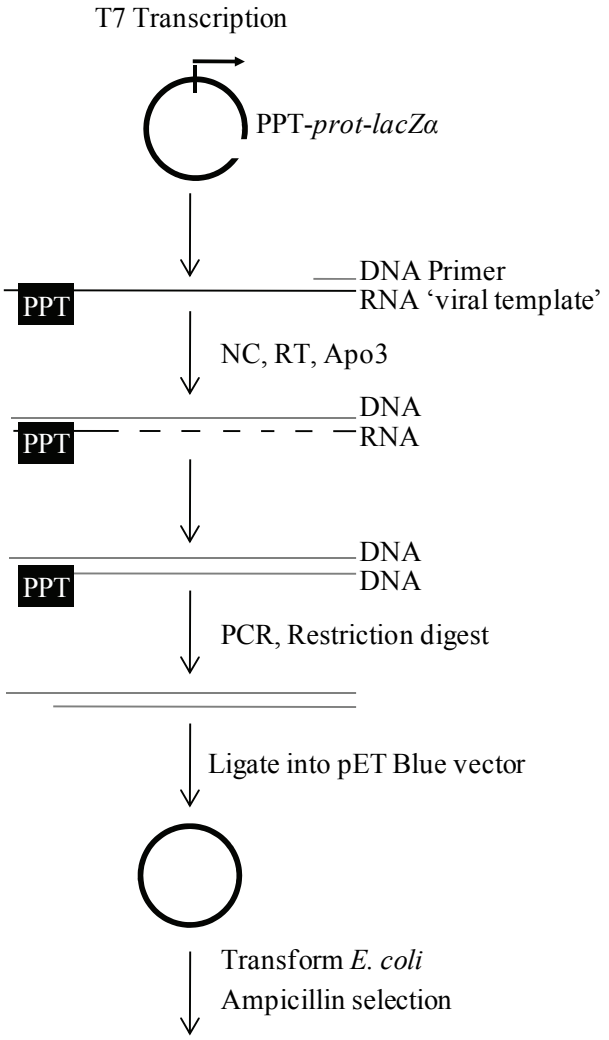


Figure 2.2 **Reconstituted HIV replication system.** Sketch depicting the steps of the reconstituted HIV replication system. The dsDNA product is meant to model a provirus and contains 120 nt of the HIV-1 *prot* gene (2282-2401 nt) and the *lacZα* (236 nt).

Table 2.2 **Primers used for HIV reverse transcription assay.**

Name	Sequence
PCR Amplification Forward Primer	GGG ACT CGA GAT AGG AGG ACA ACT GAA AGA AG
PCR Amplification Reverse Primer	GGG CAG GAA TTC GAT ATC TAT ATC
Plasmid Sequencing Primer	CGG CGA CGA CCG GTG AAT TG

2.6.2 PCR amplification of deaminated product

Once purified, the deaminated dsDNA product must be amplified to convert uracil lesions to mutations. The PCR reaction was set up as follows: deaminated product, 10 µL; dNTPs, 10 mM; Pfu C_x Buffer (Agilent Technologies), 1X; PfuTurbo C_x Hotstart DNA Polymerase (Agilent Technologies), 1U; PCR Amplification Forward primer, 200 ng; PCR Amplification Reverse primer, 200 ng. PCR reaction was adjusted with ddH_2O to a volume of 50 µL. PfuTurbo C_x can accurately use uracil as a template. The PCR cycle was as follows: 95°C for 2 min to denature the DNA; 95°C for 1 min, 50°C for 1 min, 72°C for 1 min, for 30X. Amplified fragment size was confirmed by using agarose gel electrophoresis (1% agarose). A successful amplified fragment appeared at approximately 350 nt. Once the size of the fragment was confirmed by gel, the sample was cleaned using a PCR clean up kit (Sigma) as per manufacturer's instructions.

2.6.3 Restriction digestion and ligation reactions

The purified dsDNA product was ligated into the pET-ppt-Prot vector using *Xho*I/*Eco*RI sites which enabled the mutated *prot-lacZα* to replace the *prot-lacZα* originally contained in the plasmid. The reaction conditions were as follows: Cleaned PCR product, 10 µL; *Eco*RI restriction enzyme, 1U; *Xho*I restriction enzyme, 1U; bovine serum albumin, 1X; NEB Buffer-4 (NEB), 1X. The 15 µL reaction was incubated at 37°C for 1.5 hours followed by heat inactivation at 65°C for 20 min.

Following the restriction digestion of the cleaned PCR production, we ligated it into the pET-Blue-PPT-Prot transformation vector. The reaction conditions were as follows: cleaned PCR production (double restriction digested), 2 µL; pET-Blue-1 vector (double restriction digested with *Eco*RI/*Xho*I and gel purified), 50 ng; T4 DNA ligase (NEB), 1U; and ligation buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT), 1X. The 10 µL reaction was incubated at room temperature for 30 min.

2.6.4 Transformation and plasmid purification

To screen for mini-proviruses that were deaminated during the reverse transcription reaction, they had to be transformed into *E. coli* for blue/white screening. Competent DH5α cells were made in the lab using the Z-Competent *E. coli* Transformation Kit (Zymo Research)

as per the instruction of the kit. SOB broth (20 g/L Tryptone, 5 g/L Yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 0.5 mM KCl, pH 7) was substituted for proprietary Zymobroth because competency results were similar. The heat shock method was used for transformation. Cells were plated in triplicate on LB plates containing IPTG, X-gal and Ampicillin.

White/light blue colony frequency was recorded after a minimum of 16 hours of incubation at 37°C. *E. coli* cells that received an unmutated 'provirus' formed colonies that appeared dark blue. *E. coli* cells that received a mutated 'provirus' formed colonies that appeared white or light blue. The *lacZ* reporter is very sensitive to both missense (light blue colonies) and nonsense (white colonies) mutations (Bebenek and Kunkel, 1995). Twenty-five white/light blue colonies were selected at random for sequencing. Selected colonies were each used to inoculate 5 ml LB-AMP (Ampicillin, 100 µg/mL) broth overnight in culture tubes and a streak plate for additional colour verification. Sequencing of blue colonies demonstrated an absence of significant mutations and confirmed that selection of white/light blue colonies did not bias the experimental results.

After 16 hours of incubation, a mini plasmid preparation was performed on each of the 25 selected colonies. Cultures were pelleted at 2000 *x g* for 10 min at 4°C. Supernatant was decanted and the pellet was re-suspended in 250 µL of P1 buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/mL RNase A). The mixture was then transferred to a 1.5 mL centrifuge tube and mixed by inversion (3-4 times) with P2 buffer (200 mM NaOH, 1% W/V SDS). Once the solution clarifies, 350 µL P3 (3 M Potassium Acetate pH 5.5) is added and mixed. The neutralized mixtures were then centrifuged at 15 000 *x g* for 15 min. The supernatant was then transferred to a fresh 2 mL tube. Plasmid DNA was precipitated by mixing the sample with 0.7X volume of 100% isopropanol and incubating at room temperature for 30 min. Precipitated plasmid DNA was pelleted by centrifugation at 15 000 *x g* for 20 min. Isopropanol was removed by careful aspiration without disturbing the pellet. The pellet was then washed with 350 µL 70% ethanol and centrifuged again at 15 000 *x g* for 20 min. The ethanol was removed carefully and the DNA pellet was resuspended in 25 µL ddH_2O. DNA sequencing was carried out at the National Research Council (Saskatoon, Saskatchewan).

2.7 Active transcription assay

2.7.1 Radiolabelling reaction

DNAs were designed based on Pham *et al.* (2003) (Table 2.3). The transcription reaction sequencing primer (Table 2.3) was 5'-end labeled with ^{32}P for detection. The concentrations of the reaction were as follows: Transcription Sequencing Primer, 600 nM; T4 PNK (New England Biolabs), 1U; PNK Buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, and 5 mM DTT); and γ-^{32}P ATP (70% labelling efficiency, 6000 Ci/mmol, 1.7 μM, Perkin-Elmer), 420 nM. Reaction mixture was heated to 37°C for 1 hour and heat inactivated at 65°C for 20 min.

2.7.2 Substrate annealing

The double stranded transcription substrate was formed by heat annealing the transcribed strand to the corresponding non-transcribed strand (Table 2.3). The annealing reaction concentrations were as follows: Transcribed Strand (TTC, CCC, CTC, and AGC), 360 nM; Non-Transcribed Strand (TTC, CCC, CTC, and AGC), 300 nM; Heat Annealing buffer (50 nM Tris pH 7.5, and 50 mM NaCl). The Transcribed Strand was annealed in excess to ensure that no Non-Transcribed Strand remained single stranded. The annealing reaction mixture was heated to 85°C and cooled gradually.

2.7.3 Transcription reaction

The annealed substrates were reacted with T7 RNA polymerase and Apo3 enzymes. The reaction conditions were as follows: Annealed Transcription Substrate (TTC, CCC, CTC, and AGC), 30 nM; T7 RNA Polymerase (Promega), 1U; ribonucleotides (rNTPs) (Promega), 500 μM; Apo3/AID, 30-300 nM; RNase A (DNase-Free, Roche), 200 ng; Transcription Buffer (50 mM Tris pH 7.5, 10 mM $MgCl_2$, and 1 mM DTT). The reaction mixture was heated to 37°C for 60 min with samples taken at predetermined times (0 min, 5 min, 7.5 min, 10 min, 15 min, 30 min). The reaction was stopped by mixing phenol:chloroform:isoamyl with sample. DNA was extracted using Light Phase Lock tubes (5-Prime) and washed with chloroform. The non-transcribed strand was then sequenced as described in section 2.7.4.

2.7.4 Non-transcribed strand sequencing reaction

In this experiment, deaminations are determined by sequencing the non-transcribed strand of DNA and identifying any base substitutions that have occurred. Two modified Sanger sequencing reactions were set up, one with ddA and a second with ddG. The typical sequencing reaction involves halting the reaction at a selected dideoxynucleotide. The sequencing requires that the equivalent deoxynucleotide be included in the reaction so that sequencing can continue through the potential stop. In the sequencing reactions we used, the dideoxynucleotide (ddA or ddG) represents the only adenine or guanine.

First the PCR master mix was set up as follows: γ-^{32}P ATP 5'-end labeled Transcription sequence primer (Table 2.3), 33 nM; Transcription reaction, 10 μL; Sequenase buffer (Affymetrix) (26 mM Tris-HCl pH 9.5, 6.5 mM $MgCl_2$); Thermo Sequenase DNA Polymerase (Affymetrix), 2 μL; ddH_2O to 18 μL. The PCR master mix was then pipette mixed and placed on ice. Next, 2 thin walled PCR tubes were set up. The ddG sequencing tube was set up as follows: PCR master mix, 18 μL; dC, 150μM; dA, 150 μM; dT, 150 μM; ddG, 150 μM. The ddA sequencing tube was set up as follows: PCR master mix, 18 μL; dC, 150μM; dG, 150 μM; dT, 150 μM; ddA, 150 μM. The PCR reaction was run with the following program: 94°C for 2 min to denature the DNA; 94°C for 30 sec, 55°C for 45 sec, 72°C for 1 min, 6X; 4°C once complete.

2.7.5 Gel electrophoresis and detection

DNA samples were separated on 35 cm 20% PAGE sequencing gels (20% Acrylamide:Bis-Acrylamide 19:1, 8 M urea, 1X TBE Buffer (90 mM Tris, 90 mM Boris acid, 2 mM EDTA pH 8.0) with 1X TBE running buffer. The gels were pre-run for 30 min at 60W. Samples were heat denatured for 2 min at 85°C, then placed on ice immediately prior to loading into gel. One to 3 μL of sample was loaded onto gel. Gels were run at 60W for approximately 3 hours, where a Xylene Cyanol dye marker has migrated approximately 27 cm. The gel was then transferred to paper and dried for 2 hours with a heat/vacuum gel drier (BioRad). Once dry, gels were exposed to an Imaging Screen-K (BioRad) between 0.5-12 hours depending on the activity of the sample. The phosphorimaging Screen was scanned using either a Typhoon Trio multipurpose scanner (GE Healthcare) or a Pharos FX scanner (BioRad).

Table 2.3 **Substrates and primers used for active transcription assay.**

Name	Sequence
Transcribed Strand AGC	ACC TCC TCA CCT TTC CTC TTT GCT TTC CCC TTT TCT CCC TAT AGT GAG TCG TAT TAT GCA TCT CTG
Non-transcribed Strand AGC	CAG AGA TGC ATA ATA CGA CTC ACT ATA GGG AGA AAA GGG GAA **AGC** AAA GAG GAA AGG TGA GGA GGT
Non-transcribed Strand TTC	CAG AGA TGC ATA ATA CGA CTC ACT ATA GGG AGA AAA GGG GAA **TTC** AAA GAG GAA AGG TGA GGA GGT
Transcribed Strand TTC	ACC TCC TCA CCT TTC CTC TTT GAA TTC CCC TTT TCT CCC TAT AGT GAG TCG TAT TAT GCA TCT CTG
Non-transcribed Strand CTC	CAG AGA TGC ATA ATA CGA CTC ACT ATA GGG AGA AAA GGG GAA **CTC** AAA GAG GAA AGG TGA GGA GGT
Transcribed Strand CTC	ACC TCC TCA CCT TTC CTC TTT GAG TTC CCC TTT TCT CCC TAT AGT GAG TCG TAT TAT GCA TCT CTG
Non-Transcribed Strand CCC	CAG AGA TGC ATA ATA CGA CTC ACT ATA GGG AGA AAA GGG GAA **CCC** AAA GAG GAA AGG TGA GGA GGT
Transcribed Strand CCC	ACC TCC TCA CCT TTC CTC TTT GGG TTC CCC TTT TCT CCC TAT AGT GAG TCG TAT TAT GCA TCT CTG
Transcription Sequence Primer	ACC TCC TCA CCT TTC CTC

Substrates were designed based on experiment by Pham *et al.* 2003.

CHAPTER 3.0

RESULTS

3.1 Deamination activity of Apo3A

3.1.1 Specific Activity

Previous to this research Apo3A had not been purified and studied biochemically. It was therefore important to determine if the protein that we purified retained activity and that the preferred deamination motif was similar to previous reports using virus packaged Apo3A (Chen *et al.*, 2006). The assay consists of reacting an Apo3 enzyme with a synthetic DNA that contains two staggered cytosine motifs (Figure 3.1, *sketch*). Each Apo3 enzyme has a di- or tri-nucleotide sequence in which it catalyzes deaminations with the greatest specific activity. For Apo3A, the preferred dinucleotide sequence has been determined to be 5'TC (Chen *et al.*, 2006). The minimal motif 5'TC is conserved among most Apo3 enzymes except for Apo3G which has a strong preference for 5'CC<u>C</u> (underlined C is preferentially deaminated) (Beale *et al.*, 2004; Suspene *et al.*, 2004; Yu *et al.*, 2004). Substrates were analyzed *in silico* to ensure they would not form significant secondary structure. Apo3A was found to deaminate at 5'TC sites (Table 2.1) and we extended this observation to determine whether Apo3A has a conservative or more promiscuous target motif. We determined the specific activity of the enzyme on ssDNA substrates (85-nt) containing 5'TTC, 5'ATC, 5'CTC, 5'GTC, or 5'ACC motifs embedded within the flanking 5' AAA NNN ATA that have been shown as optimal for Apo3G activity (Table 2.1) (Chelico *et al.*, 2006). Apo3A appears to deaminate all 5'TC containing sequences about equally and did not preferentially deaminate motifs with a specific 5'-base by more than 2-fold (Table 3.1, *specific activities of 3.1 to 6.1 pmol/µg/min*). However, there is a 100-fold decrease in activity when the 5'ACC sequence was used, suggesting that the dinucleotide 5'TC is necessary for optimal deamination activity (Table 3.1), in agreement with a previous report (Chen *et al.*, 2006). Apo3A was unable to deaminate a double stranded substrate (Table 3.1). In comparison, Apo3G does not have a promiscuous target motif and strictly deaminates the third C in a CCC motif (Beale *et al.*, 2004; Suspene *et al.*, 2004; Yu *et al.*, 2004).

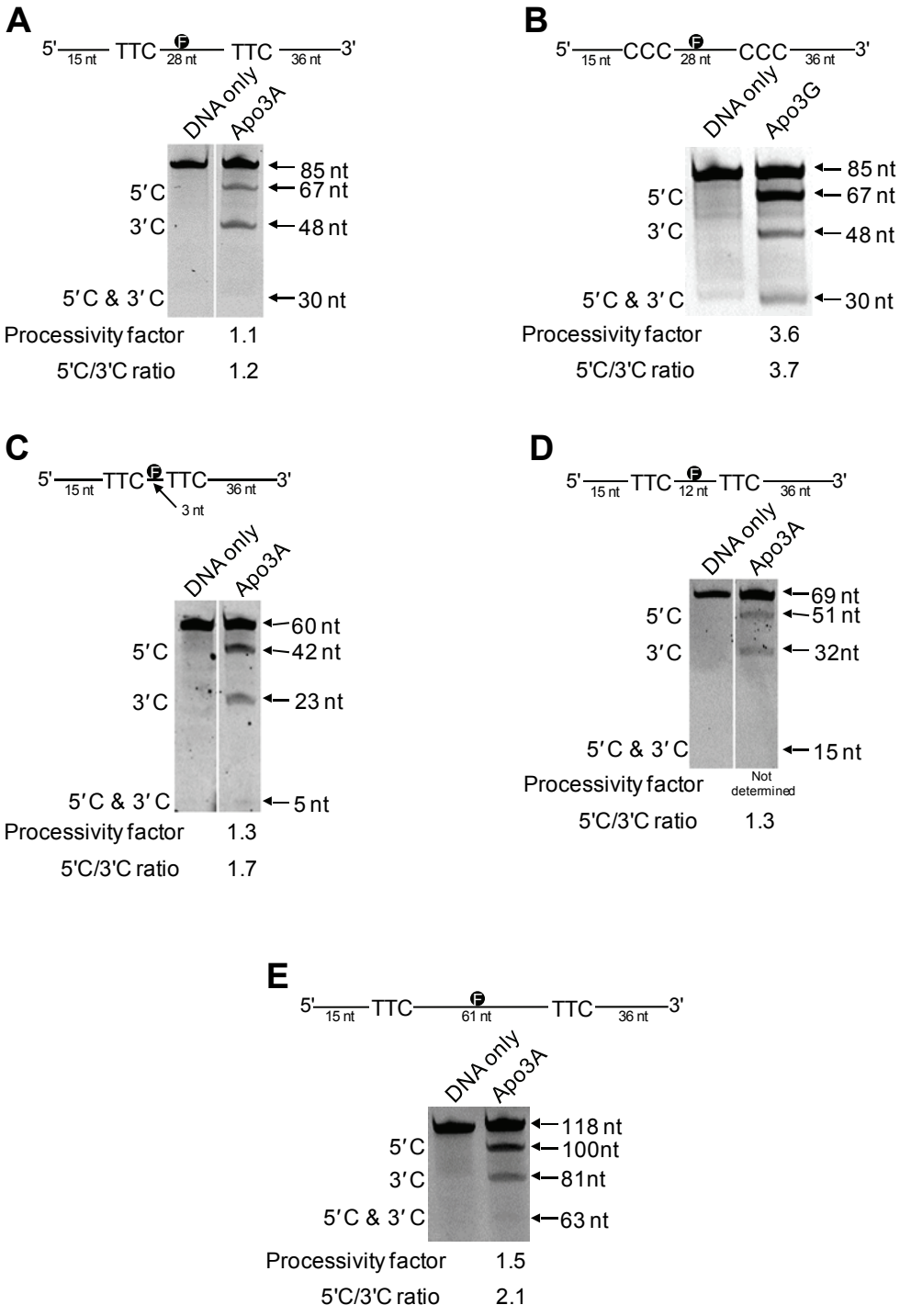

A
5'
15 nt
TTC
F
28 nt
TTC
36 nt
3'
DNA only
Apo3A
85 nt
67 nt
48 nt
30 nt
5′C
3′C
5′C & 3′C
Processivity factor
1.1
5'C/3'C ratio
1.2
B
5'
15 nt
CCC
F
28 nt
CCC
36 nt
3'
DNA only
Apo3G
85 nt
67 nt
48 nt
30 nt
5′C
3′C
5′C & 3′C
Processivity factor
3.6
5'C/3'C ratio
3.7
C
5'
15 nt
TTC
F
TTC
36 nt
3'
3 nt
DNA only
Apo3A
60 nt
42 nt
23 nt
5 nt
5′C
3′C
5′C & 3′C
Processivity factor
1.3
5'C/3'C ratio
1.7
D
5'
15 nt
TTC
F
12 nt
TTC
36 nt
3'
DNA only
Apo3A
69 nt
51 nt
32nt
15 nt
5′C
3′C
5′C & 3′C
Processivity factor
Not determined
5'C/3'C ratio
1.3
E
5'
15 nt
TTC
F
61 nt
TTC
36 nt
3'
DNA only
Apo3A
118 nt
100nt
81nt
63 nt
5′C
3′C
5′C & 3′C
Processivity factor
1.5
5'C/3'C ratio
2.1

Figure 3.1 **Analysis of Apo3A processivity on ssDNA.** A-B, deamination experiment using a Fam-labeled 85 nt synthetic substrate that is specific for Apo3A (A) and Apo3G (B). Deaminations are detected as three bands caused by breakage at C→U lesions processed during the experiment. A 5' deamination results in a 67 nt fragment, a 3' deamination results in a 48 nt fragment while a deamination at both motifs results in a 30 nt fragment. C, Apo3A catalyzed deamination of a 60 nt Fam-labeled DNA substrate. Deaminations result in a 5' band at 42 nt, 3' band at 23 nt, and a double deamination band at 5 nt. D, Apo3A catalyzed deamination of a 69 nt Fam-labeled DNA substrate. Deaminations result in a 5' band at 51 nt, 3' band at 32 nt, and a double deamination band at 15 nt. E, Apo3A catalyzed deamination of a 118 nt Fam-labeled DNA substrate. Deaminations result in a 5' band at 100 nt, 3' band at 81 nt, and double deamination band at 63 nt. Values of both processivity and 5'/3' ratio are shown below gels and were determined by analysis of gel band intensities. The migration distances of the deamination bands were determined by allowing a reference reaction to deaminate substrate to completion. Sketches above gel pictures depict DNA substrates utilized. Results are averaged from at least two independent trials. The standard deviation (A and B) or the range (C and E) for the processivity factors are: 0.25 (A), 0.09 (B), 0.46 (C), and 0.57 (E). For (D), a deamination at both motifs could not be detected and precluded the calculation of a processivity factor. Figure is adapted from Love *et al.* 2012.

Table 3.1 **Specific activities of Apo3A, Apo3H, and Apo3G on ssDNA substrates.** Specific activities were determined by measuring pmoles of substrate deaminated per min for a microgram of enzyme. Values are shown ± the SD determined from three (Apo3A, Apo3G) or the range of 2 (Apo3H) independent experiments. The ds indicates double-stranded DNA.

Enzyme	Sequence	Specific Activity (pmol/min/μg)
Apo3A	aaa ACC ata	0.07 ± 0.01
	aaa ATC ata	4.6 ± 1.5
	aaa TTC ata	6.1 ± 1.9
	aaa CTC ata	4.1 ± 2.5
	aaa GTC ata	3.1 ± 2.5
	(ds) aaa TTC ata	Undetectable
Apo3H	aaa CTC ata	8.2 ± 1.0
	aaa ATC ata	1.9 ± 0.3
	aaa TTC ata	2.4 ± 1.0
	aaa GTC ata	2.1 ± 1.6
	aaa ACC ata	Undetectable
Apo3G	aaa CCC ata	21.3 ± 0.9

3.1.2 Processivity

Apo3G was shown to deaminate cytosines processively using a 3-dimensional search that is characteristic of facilitated diffusion (Chelico *et al.*, 2006; Feng and Chelico, 2011; Halford and Marko, 2004). In this type of processive movement, the enzyme can slide and make microscopic dissociations and reassociations with the DNA (termed hopping, jumping or intersegmental transfer), without diffusing into the bulk solution, in order to efficiently search for target sequences where catalysis takes place (Chelico *et al.*, 2006; Feng and Chelico, 2011; Halford and Marko, 2004). Apo3G has also been shown to prefer deamination towards the 5'-end of linear ssDNA (Figure 3.1 B) (Chelico *et al.*, 2010; Chelico *et al.*, 2008; Senavirathne *et al.*, 2012). Apo3G can move on ssDNA in both the 5' and 3' direction, but can only catalyze deamination approaching a 5'CCC motif from 3' to 5'. This has been termed a catalytic orientation specificity (Chelico *et al.*, 2006). It is not known if these hallmark features of Apo3G are unique or found in other Apo3 family members. Here Apo3G is used as a control for Apo3 activity as many of its biochemical characteristics have been determined. As will become clear throughout the presentation of our work, Apo3G is unique within the Apo3 family.

Substrates used to determine processive characteristics of Apo3A contain two deamination motifs separated by a given distance and contain a Fam-labeled thymine in between the two deamination motifs. The internal Fam-label allows identification of deaminations that occurred at both 5'TTC motifs on the same ssDNA substrate. All deamination reactions were performed under 'single hit' conditions (<15% substrate usage) (Creighton *et al.*, 1995). Single hit conditions were determined from research performed with polymerases that determined that when less than 15% of a given template is extended, each of the extended products is the result of a single interacting polymerase with the substrate (Creighton *et al.*, 1995). As the percentage of templates extended increases, the likelihood that more than one polymerase catalyzed the extension reaction is also increased.

To study Apo3A, we used a substrate with two 5'TTC motifs separated by 28-nt (Figure 3.1 A, *sketch*). The specific activities for Apo3A and Apo3G on these substrates were 6.1 pmol/min/μg and 21.3 pmol/min/μg, respectively (Table 3.1). On this substrate, Apo3A did not catalyze the deamination of both sites processively as seen with the processivity factor of 1.1 (Figure 3.1 A, *below gel*). The processivity factor is a ratio of the observed double deaminations

(Figure 3.1 A, *5'C & 3'C band*) to the calculated expected value if the double deaminations were from independent events of two enzymes, i.e., not processive. Therefore, a processivity factor of 1 indicates that the two deaminations on the substrate occurred by independent events. Importantly, if deaminations were allowed to take place outside of single-hit conditions (>15% substrate usage), then double deamination events were clearly seen on the gel indicating that the substrate can be acted upon by Apo3A in two independent events and be deaminated to near completion (data not shown). Apo3A also does not show a preference for deamination of either the 3'- or 5'-proximal motif (Figure 3.1 A, *5'C/3'C ratio is 1.2*). The processive enzyme Apo3G demonstrates very different deamination characteristics on its cognate substrate (Figure 3.1 B). Apo3G has a processivity factor of 3.6 (Figure 3.1*B*, *below gel*). This means that Apo3G is ~4-fold more likely to catalyze a processive deamination of two 5'CCC motifs than deaminate each motif in a separate enzyme-substrate encounter. Apo3G also shows a characteristic preference for deamination of cytosine motifs at the 5'-end (Figure 3.1 B, *5'C/3'C ratio is 3.7-fold*) (Chelico *et al.*, 2006).

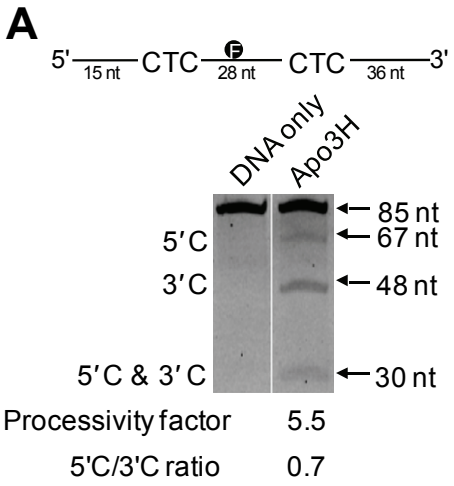
To determine whether Apo3A could deaminate C residues processively if the distance between the 5'TTC motifs was changed, we tested substrates that had either a 3- or 12-nt distance between deamination motifs (Figure 3.1 C-D). For the substrates with the 5'TTC motifs 3-nt apart, Apo3A had a low processivity (Figure 3.1 C, *processivity factor 1.3*), suggesting that minimal local scanning can take place. However, when the substrate had the 5'TTC motifs 12-nt apart, Apo3A did not catalyze processive deaminations as demonstrated by no double deamination band being detectable under single-hit conditions, which precluded the calculation of a processivity factor (Figure 3.1 D, *absence of 5'C & 3'C band*). To ensure that a double deamination was possible for the 12 nt substrate, Apo3A was reacted with the substrate until a band could be visualized (>15% substrate usage) (data not shown). Based on data from a mutant Apo3G that prefers to jump more than slide, the jumping movement appears to favor deamination motifs that are spaced further apart (Feng and Chelico, 2011). On a substrate with the 5'TTC motifs separated by 61 nt, Apo3A demonstrates a processivity factor of 1.5 and a polarity preference of 2.1 for the 5'-proximal C (Figure 3.1 E). That the processivity factor is above 1 for this substrate indicates that Apo3A may be able to catalyze processive deaminations of distantly spaced cytosines (>30 nt) a small proportion of the time (Figure 3.1 E). All together the data (Figure 3.1 A, C-D) indicate that Apo3A has a limited ability for local sliding (Figure

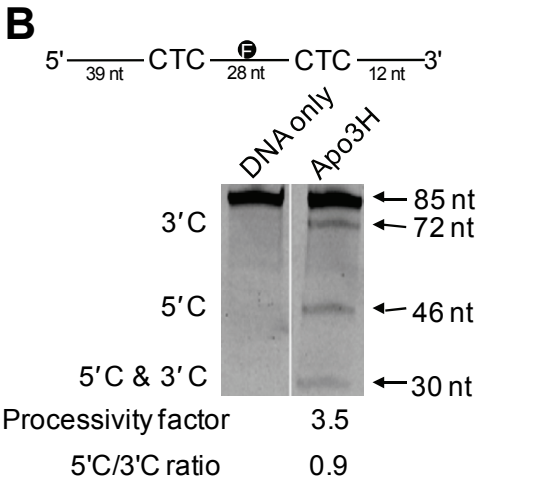
3.1 C) and distal jumping (Figure 3.1 E), suggesting that the large majority of deaminations catalyzed by Apo3A are non-processive.

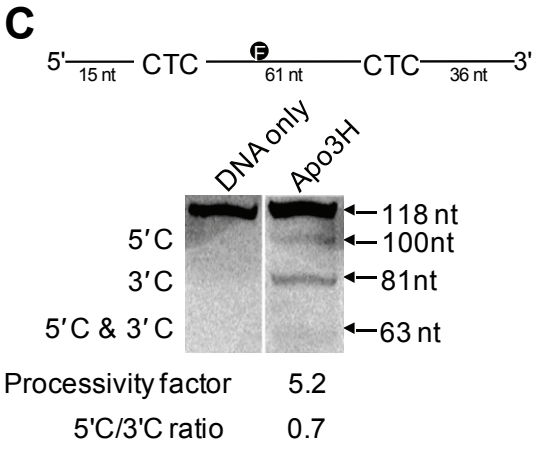
3.2 Deamination activity of Apo3H

To address the evolutionary pressure on Apo3H that had caused its inactivation, we developed two hypotheses. Either the Apo3H activity had become too great and was resulting in cellular DNA damage, or that the activity had been lost to the point where the enzyme was essentially useless against retroelement restriction. We used a biochemical approach to determine the inherent properties of the enzyme, which could provide evidence for the "incorrect" function or loss of function of Apo3H in comparison to other APOBEC enzymes.

We used the deamination assay to determine if the activity of Apo3H was different from other Apo3 family members. The system consisted of an 85 nt substrate with staggered cytosine motifs (Figure 3.2 A, *sketch*). A variety of substrates were used to determine the motif with the highest specific activity (Table 3.1). The 5'CTC motif resulted in the highest activity and was used in all subsequent experiments (Table 3.1). The activity of Apo3H (8.2 pmol/µg/min) was 25% higher than Apo3A (6.1 pmol/µg/min), but lower than Apo3G (21.3 pmol/µg/min) (Table 3.1). This addressed the initial hypothesis that evolutionary pressure on Apo3H may be due in part to its proficiency in catalyzing deaminations. Second, the first deamination experiment demonstrated that Apo3H was able to catalyze deaminations in a processive manner (Figure 3.2 A, *below gel*). Previously it has been demonstrated that the non-catalytic domain of Apo3G was necessary for processivity (Chelico *et al.*, 2010). This is thought to be necessary because the Apo3G catalytic domain is negatively charged and requires the positively charged N-terminal domain to stay associated with ssDNA (Chelico *et al.*, 2010). Apo3H does not possess a second inactive domain to mediate processivity as does Apo3G (Chelico *et al.*, 2010). It has however, a high positive charge that may allow it to bind tightly to DNA (addressed in Section 3.4).


A
5'
15 nt
CTC
F
28 nt
CTC
36 nt
3'
DNA only
Apo3H
85 nt
67 nt
5′C
3′C
48 nt
5′C & 3′C
30 nt
Processivity factor
5.5
5'C/3'C ratio
0.7


B
5'
39 nt
CTC
F
28 nt
CTC
12 nt
3'
DNA only
Apo3H
85 nt
3′C
72 nt
5′C
46 nt
5′C & 3′C
30 nt
Processivity factor
3.5
5'C/3'C ratio
0.9


C
5'
15 nt
CTC
F
61 nt
CTC
36 nt
3'
DNA only
Apo3H
118 nt
5′C
100nt
3′C
81nt
5′C & 3′C
63 nt
Processivity factor
5.2
5'C/3'C ratio
0.7

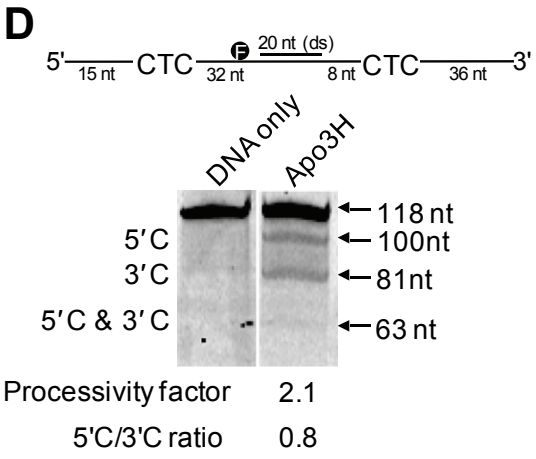

D
5'
15 nt
CTC
F
32 nt
20 nt (ds)
8 nt
CTC
36 nt
3'
DNA only
Apo3H
118 nt
5′C
100nt
3′C
81nt
5′C & 3′C
63 nt
Processivity factor
2.1
5'C/3'C ratio
0.8

Figure 3.2 **Analysis of Apo3H processivity on ssDNA.** A, deamination experiment using a Fam-labeled 85 nt DNA synthetic substrate that is specific for Apo3H (CTC motif). Deaminations are detected as three bands caused by breakage at C→U lesions catalyzed during the experiment. A 5'C deamination results in a 67 nt fragment, a 3'C deamination results in a 48 nt fragment while a deamination at both motifs results in a 30 nt fragment. B, Apo3H catalyzed deamination of an 85 nt Fam-labeled DNA substrate. Deaminations results in a 3'C band at 72 nt, 5'C band at 46 nt, and double deamination band at 30 nt. C, Apo3H catalyzed deamination of a 118 nt Fam-labeled DNA substrate. Deaminations result in a 5'C band at 100 nt, 3'C band at 81 nt, and double deamination band at 63 nt. D, Apo3H catalyzed deamination of a 118 nt Fam-labeled DNA substrate as in C, but with a 20 nt DNA spacer (blue) annealed between cytosine motifs. Values of both processivity and 5'C/3'C ratio are shown below gels and were determined by analysis of gel band intensities. The migration distances of the deamination bands were determined by allowing a reference reaction to deaminate substrate to completion. Sketches above gel pictures depict DNA substrates utilized. DNA samples were heat denatured (2 min at 85°C) prior to loading and were separated under denaturing conditions. Results are averaged from at least two independent trials. The standard deviation (A) or range (B, C, and D) of processivity values are: 0.42 (A), 0.01 (B), 2.55 (C), and 0.28 (D).

3.2.1 Polarity

Apo3H also demonstrated a notable bias towards the 3'C motif of the DNA (Figure 3.2 A, *5'C/3'C ratio 0.7*). This stood out as uncharacteristic for an Apo3 enzyme based on the data obtained for Apo3G which shows a strong 5'C motif bias (Figure 3.1 B, *5'C/3'C ratio 3.7*), as well as the data obtained for Apo3A where there was little 5'-bias (Figure 3.1 A, *5'C/3'C ratio 1.2*). We investigated the directional bias of Apo3H by creating a substrate that had the cytosine motifs in a reversed orientation (Figure 3.2 A and B, *compare substrate sketches*). In this way, we could see either that the bias towards the 3' end of the DNA was retained, or that the directional bias was due to a preference of the enzyme to deaminate motifs flanked by a greater number of nucleotides i.e., motifs closer to the centre of the ssDNA. It was found that Apo3H retained its preference for the 3'-end of the ssDNA, but that because the motif had been moved closer to the centre of the DNA there was a increase in 5'C deamination (Figure 3.2 A-B, *compare 5'C/3'C ratios 0.7 to 0.9*). These data suggest that Apo3H can catalyze a deamination reaction if there is a greater amount of ssDNA flanking the 5'-end of the motif. This appears to be the opposite of Apo3G which required more ssDNA flanking the 3'-end of the motif (Chelico *et al.*, 2008).

3.2.2 Processivity

We investigated the ability of Apo3H to catalyze processive deaminations using various custom substrates. On a substrate with the CTC motifs separated by 28 nt, Apo3H demonstrated the strongest processivity (Figure 3.2 A, *below gel, processivity factor 5.5*) in comparison to Apo3A (Figure 3.1 A, *below gel, processivity factor 1.1*) or Apo3G (Figure 3.1 B, *below gel, processivity factor 3.6*). We tried increasing the distance between the cytosine motifs to 61 nt and observed a processivity factor of 5.2 (Figure 3.2 C). Previous research using Apo3G processivity mutants has demonstrated that increasing the distance between two cytosine motifs for an enzyme that prefers to slide results in a reduction in the processivity (Chelico *et al.*, 2006; Feng and Chelico, 2011). This is because Apo3G uses two different mechanisms for scanning DNA, both a sliding motion where Apo3G scans every nucleotide and a jumping motion in which microscopic dissociations allow translocations of >20 nt (Feng and Chelico, 2011). In the case of Apo3H, the jumping motion is likely part of the method of scanning employed because no loss in processivity is observed. To determine whether sliding was also involved, we

challenged Apo3H processivity to observe if it was reduced by an experimental block of any potential sliding movements of Apo3H. We introduced a DNA block that annealed to the substrate DNA between the cytosine motifs (Figure 3.2 D, *sketch*), which has been shown to block the sliding component of Apo3G (Chelico *et al.*, 2006). We found that there was ~2-fold reduction in the processivity when the dsDNA block was annealed (Figure 3.2 C, *compare 2.1 from 5.2*), indicating that Apo3H does use sliding for scanning DNA in combination with jumping. To quantify the contribution of sliding versus jumping we compared the ratio of successful coordinated deaminations i.e. if Apo3H deaminates the 5'-end motif and proceeds to deaminate the 3'-end motif in a single interaction for both the fully ssDNA substrate (Figure 3.2 C) and the partially dsDNA substrate (Figure 3.2 D). To do this, we compared the ratio of (5'C&3'C)/5'C deaminations of fully ssDNA and found that Apo3H successfully deaminates both C targets at a frequency of 0.89 (Figure 3.2 C). When a DNA block is introduced the ratio of successful processive deaminations is reduced to a frequency of 0.13 as calculated by the (5'C&3'C)/5'C ratio. This means that jumping contributes to 15% of the total processive movement (0.13/0.89) and sliding contributes to 85% of the movement. In contrast Apo3G has been found to use jumping 60% of the time and sliding 40% of the time (Feng and Chelico, 2011).

3.3 Apo3A weakly binds ssDNA

Based on the deamination data (Figure 3.1 A, C-E) that Apo3A was not able to catalyze processive deaminations with great efficiency, we hypothesized that this could be due to poor DNA binding properties. We selected Apo3G as a control based on its ability to deaminate processively (Figure 3.1B and (Chelico *et al.*, 2006)) as well as reports found in the literature that have measured its affinity for ssDNA (Chelico *et al.*, 2006; Iwatani *et al.*, 2006).

Rotational anisotropy was used to determine the apparent dissociation constant (K_d) of Apo3A, Apo3H, and Apo3G to a single stranded DNA substrate. Rotational anisotropy is the measure of the change in polarization of emitted light from a fluorophore that has been excited by vertically polarized light. The change in polarization is dependent on the rotational speed of the labeled molecule. Molecules with increased mass tumble more slowly than molecules with less mass. Therefore if a binding partner interacts with the fluorescently labeled molecule e.g. DNA, there will be a decrease in the change of the polarization state.

It was determined that Apo3A had an apparent K_d of 9.2 μM compared to 0.14 μM for Apo3G (Figure 3.3). Apo3A was not able to effectively bind to ssDNA, likely due to its predicted charge of -2.5 at a pH of 7.4. Both Apo3A and Apo3G binding curves were fit to a Hill sigmoidal curve indicating that cooperative binding or oligomerization may be occurring (Figure 3.3). This indicates that Apo3G binds to DNA 65-fold tighter than Apo3A. These findings suggest that the inability of Apo3A to catalyze processive deaminations is in part due to its inability to bind tightly to DNA. To better characterize the interaction of Apo3H with DNA, we sought to determine its apparent K_d when binding ssDNA by rotational anisotropy. Binding of Apo3H to ssDNA demonstrated a K_d of 0.19 μM and fit a Hill sigmoidal curve (Figure 3.3 B), which is similar to Apo3G (apparent K_d of 0.14 μM, Figure 3.3 C). The deamination data obtained previously (Figures 3.1 B and 3.2 A) showed approximately a 1.5-fold greater processivity for Apo3H compared to Apo3G, however the apparent K_d values are similar (Figure 3.3 B-C). It may indicate that ssDNA binding strength alone is not the only indicator of processivity. There are likely other variables that determine processivity based on amino acids that contact the DNA, which is likely to differ between Apo3G and Apo3H that are Z2 and Z3 deaminases, respectively (Figure 1.3).

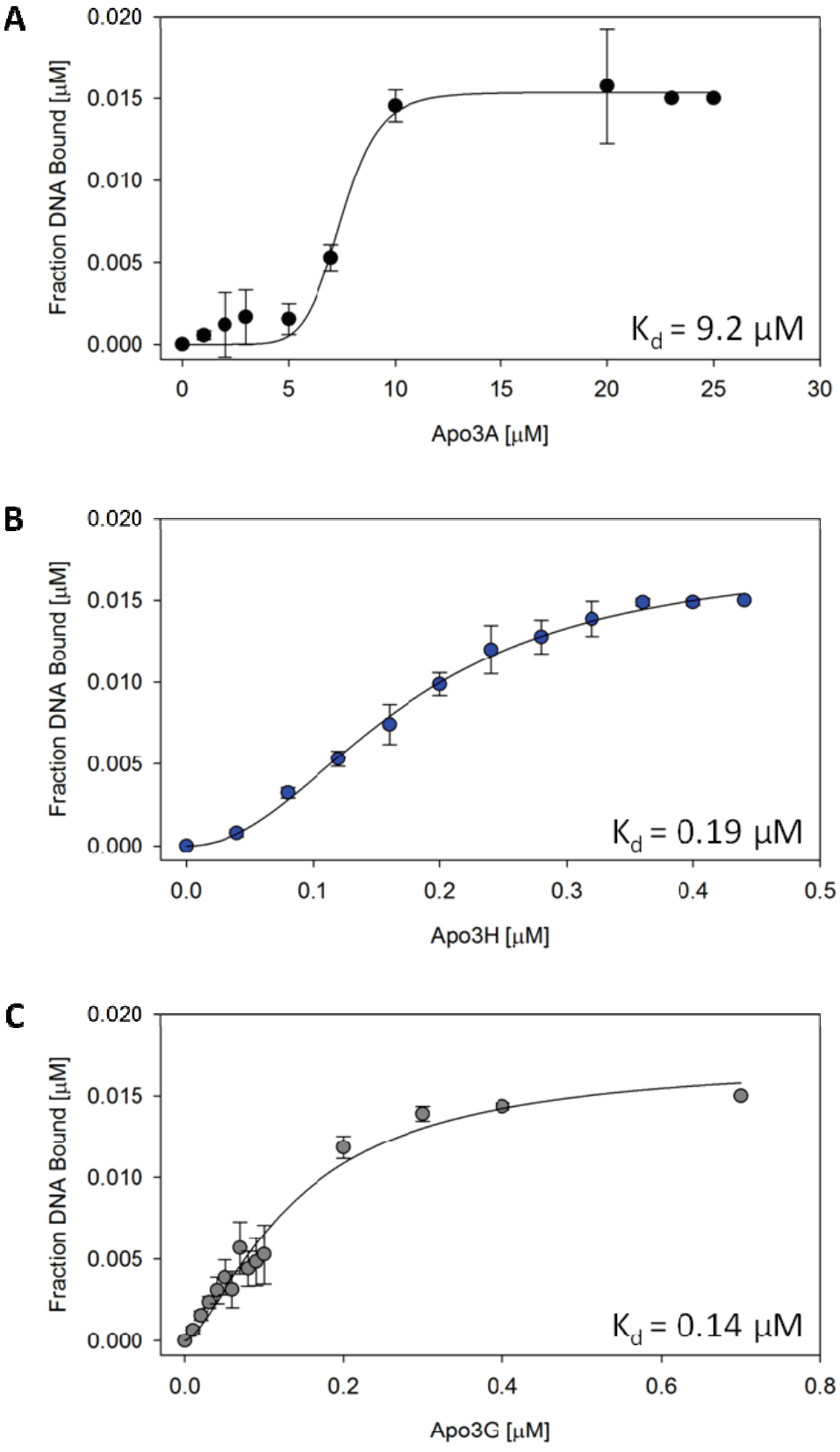

A
0.020
0.015
0.010
0.005
0.000
Fraction DNA Bound [µM]
K_d = 9.2 µM
0
5
10
15
20
25
30
Apo3A [µM]
B
0.020
0.015
0.010
0.005
0.000
Fraction DNA Bound [µM]
K_d = 0.19 µM
0.0
0.1
0.2
0.3
0.4
0.5
Apo3H [µM]
C
0.020
0.015
0.010
0.005
0.000
Fraction DNA Bound [µM]
K_d = 0.14 µM
0.0
0.2
0.4
0.6
0.8
Apo3G [µM]

Figure 3.3 **Apparent dissociation constant of Apo3A, Apo3H, and Apo3G for ssDNA.** Binding of Apo3A (A), Apo3H (B) or Apo3G (C) to ssDNA was determined using fluorescence depolarization (rotational anisotropy). Error bars represent the standard deviation of three independent experiments. The standard deviations of the K_d are (A) ± 2.5 μM, (B) ± 0.02 μM, and (C) ± 0.02. Figure is adapted from Love *et al.* 2012 (Love *et al.*, 2012).

3.4 Oligomerization state of Apo3A and Apo3G

Apo3A rotational anisotropy ssDNA binding data was best fit to a sigmoid regression curve. A sigmoidal relationship indicates that there is cooperative binding of the substrate. We used multi-angle light scattering (MALS) to determine the molecular weight and therefore the oligomerization states of Apo3A and Apo3G. MALS involves in-line analysis of protein eluted from a size exclusion column by 18 lasers at different angles and a refractive index detector. The amount of light scattering is measured and the resulting Raleigh ratio can be used to determine the absolute molecular mass of the molecule (Wyatt, 1993). Apo3A was found to be a monomer (25 kDa) in 97% of the elution. In 3% of the elution there is a distinct dimer peak (53 kDa). The MALS demonstrates that oligomers can occur (Figure 3.5 A, *3%*), but are not favorable in solution. This is consistent with the binding analysis of Apo3A (Figure 3.5 A) which suggested that oligomerization (cooperation) could be associated with DNA binding.

Apo3G was analyzed in sequence with Apo3A but demonstrated more diversity in oligomerization states. Apo3G was found to mostly elute as dimers (41%, 126 kDa) and tetramers (47%, 284 kDa). Monomeric Apo3G was found in 11% of the elution (46 kDa).

We also wished to test the oligomerization state of Apo3H but we were unable to reach the concentrations of enzyme required for the assay, approximately 1 mg/mL. We are currently in the process of optimizing the use of a cation exchange column to increase the concentration of Apo3H for MALS.

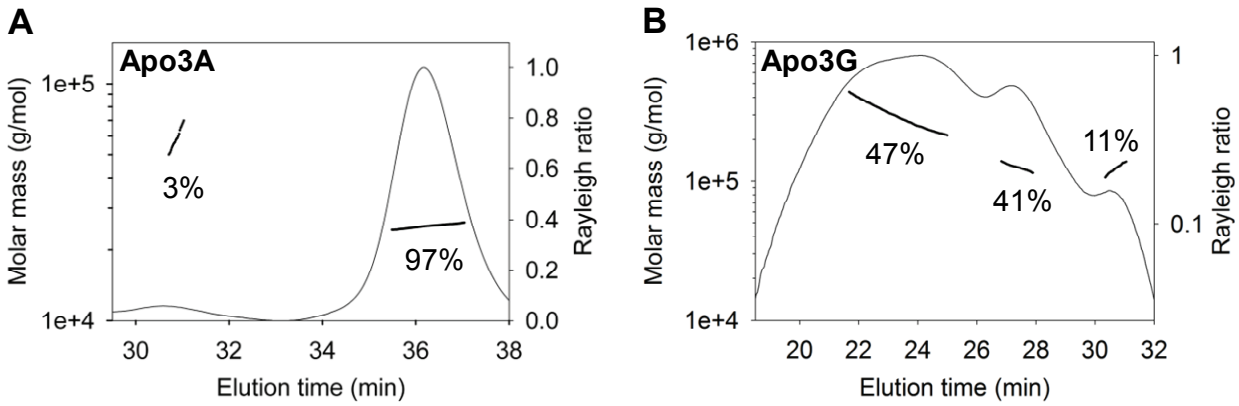


Figure 3.4 **Determination of molecular mass of Apo3A and Apo3G using multi-angle light scattering.** Purified Apo3A was resolved by size-exclusion chromatography in running buffer with 150 mM NaCl and 50 mM HEPES pH 7.3. The molecular mass is plotted throughout the eluted peaks (line plot, *left y-axis*). The Rayleigh light scattering chromatogram shows the protein distribution and is plotted on the right y-axis. The median molecular masses and distributions for Apo3A are 97% monomers (25 kDa) and 3% dimers (53 kDa). The median molecular masses and distributions for Apo3G are 47% tetramers (284 kDa), and 41% dimers (126 kDa), and 11% monomers (46 kDa).

3.5 Efficiency of HIV restriction

3.5.1 Mutational spectra induced by Apo3A, Apo3H, and Apo3G

The (-)DNA generated by RT during, HIV reverse transcription, is a target for Apo3-deamination (Yu *et al.*, 2004). It has been demonstrated that Δ*vif* HIV is unable to produce infectious virions after exiting a cell line that is expressing Apo3G (Sheehy *et al.*, 2002). The role that Apo3A and Apo3H play in the restriction of HIV has not been definitively explained. Apo3A has been shown to have strong restriction capability on HIV in myeloid cells (Berger *et al.*, 2011), but the mechanism is not known (Peng *et al.*, 2007). Apo3H has also been demonstrated to have HIV restriction ability in some studies (Harari *et al.*, 2009; OhAinle *et al.*, 2008; Tan *et al.*, 2009), and has partial resistance to Vif-mediated degradation (unpublished data) (OhAinle *et al.*, 2008; Tan *et al.*, 2009). To determine the potential role of Apo3A and Apo3H in restriction of HIV replication, we used a reconstituted HIV reverse transcription system (See Materials and Methods). For comparison, Apo3G was also tested in this system.

The reverse transcription system used here consisted of a 368 nt synthetic RNA that contained the following keys features: a primer to initiate (-)DNA synthesis; PPT, necessary for priming the elongation of the (+)DNA of the model provirus; a *lacZα* reporter gene, to determine mutated clones from those sustaining no mutations and; 120 nt of the *prot* gene from HIV-1, because it is well characterized and a complete mutagenesis study (Loeb *et al.*, 1989) can be used for reference when analyzing the consequences of Apo3 deaminations on the coding sequence. The reaction is started by the addition of dNTPs to a reaction mixture containing HIV-1 reverse transcriptase, nucleocapsid, and the synthetic RNA genome in the presence or absence of an Apo3 enzyme. An important feature of the system is the PPT which enables (+)DNA synthesis to occur as in HIV proviral DNA synthesis or HIV replication. The (+)DNA synthesis can only begin after the RNaseH domain of RT degrades the viral genomic RNA, therefore regions furthest from the PPT (the LacZα region) are single stranded the longest time and expected to incur more deaminations.

Based on Apo3G literature, processivity has been predicted to play a strong role in its ability to catalyze large numbers of deaminations because partially processive Apo3G mutants induced fewer mutations during reverse transcription (Feng and Chelico, 2011). However this

has not been assessed by comparison to a non-processive enzyme. From the results of the deamination assays discussed earlier, we knew that Apo3A was weakly processive (Figure 3.1 A, C-E). We predicted that we would see many white colonies as a result of the high amount of dissociation and reassociations events of Apo3A, but that each of the clones would carry fewer mutations when compared to Apo3G. For Apo3G we expected many white colonies due to a bimodal off rate from DNA involving short and long binders (Chelico *et al.*, 2008) with each clone containing many mutations due to processive scanning. We found that the frequency of white/light blue colonies was 0.86 for Apo3A and 0.90 for Apo3G. This number was approximately 9-fold greater than the level of mutation induced by the reverse transcription system in the absence of an Apo3 enzyme (0.10).

To compare the mutations induced by Apo3A, Apo3H, and Apo3G a mutational spectra was used (Figure 3.5). From the mutational spectra data observed for Apo3A and Apo3G, it can be seen that there were a high number of mutations induced by each enzyme (Figure 3.5) and processivity may not be essential. Apo3H was able to catalyze deaminations at a similar number of sites (Apo3H, 44 sites) when compared to Apo3A (44 sites) and Apo3G (41 sites), but the most highly mutated sites were lower for Apo3H (compare Figure 3.5 A-C, *most mutated site* Apo3A 88%, Apo3H 60%, Apo3G 92%). Upon analysis of the spectrum of mutations induced by Apo3H, it can be seen that the highest percentage of "hot spots", highly deaminated sites, are seen in the *lacZα* region. The ratio of intensities between the most deaminated site in the *lacZα* and the most deaminated site in the *prot* is 1.5 for Apo3H, while the ratio for Apo3A is 1.7 and Apo3G is 1.9. These data indicate each specific activity and processivity play a role in Apo3 deamination kinetics and that a strong ability in one of the attributes may be able to compensate for another.

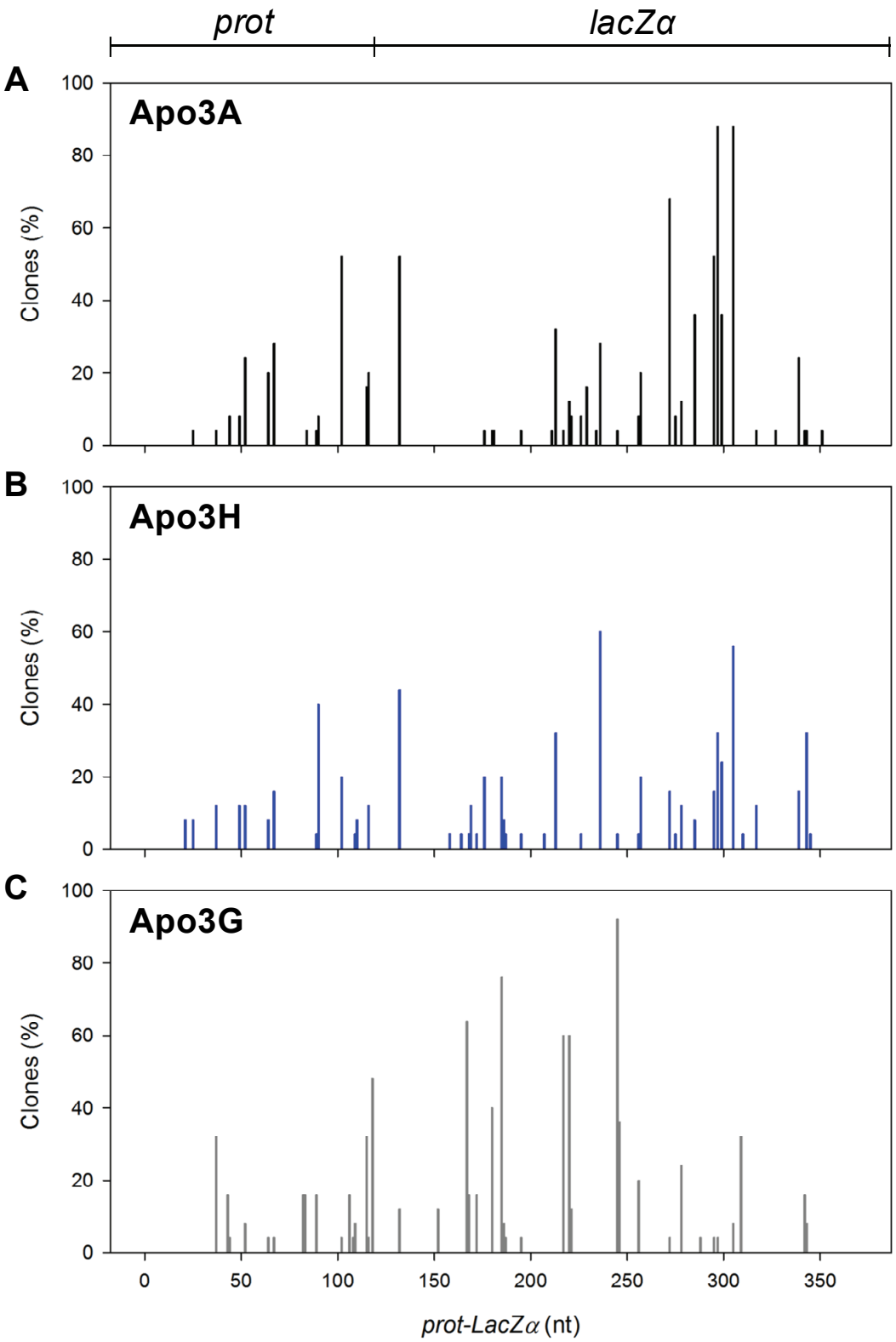

prot
lacZα
A
Apo3A
B
Apo3H
C
Apo3G
Clones (%)
0
20
40
60
80
100
0
50
100
150
200
250
300
350
prot-LacZα (nt)

Figure 3.5 **Apo3A-, Apo3H-, and Apo3G-induced mutational spectra.** Spectra of G→A mutations resulting from Apo3A- (A), Apo3H- (B), and Apo3G-catalyzed (C) deaminations occurring during reverse transcription by HIV-1 reverse transcriptase. The DNA substrate formed contains 120 nt of the HIV-1 protease (*prot*) gene (nt 2282-2401) and 248 nt of the *lacZα* reporter sequence. Data is represented as the percentage of clones containing the mutation at the indicated position. The x-axis denotes the position of the mutation in the construct. Figure is adapted from Love *et al.* 2012.

3.5.2 Mutations per clone induced by Apo3A, Apo3H, and Apo3G

The spectra results discussed thus far only reveal part of the differences between these three Apo3 proteins. We further categorized the mutated sequences selected from the reconstituted HIV reverse transcription assay based on the different numbers of mutation, sustained on a per clone basis. It was seen that in the *prot* sequence, Apo3G was able to mutate clones the greatest number of times with 28% of mutated clones having greater than 3 mutations compared to Apo3A and Apo3H which had only 16% and 8% of clones with greater than 3 mutations, respectively (Figure 3.6 A-B). The kinetics of reverse transcription is such that the substrate for Apo3-mediated deamination is revealed by RNaseH activity randomly but the time that ssDNA regions are vulnerable to deamination depends on the proximity to the PPT (See Section 1.2 and Figure 1.2). Since the *prot* region is adjacent to the PPT (Figure 2.2), it is single-stranded for the shortest time relative to the *lacZα* region.

The difference between clones generated by Apo3A, Apo3H, and Apo3G are less pronounced in the *lacZα* region. This region is available for a longer period of time for deamination, so a greater number of mutations are expected for all Apo3 enzymes. Most clones produced by the Apo3 enzymes contained 4-5 mutations (Apo3G 24%, Apo3A 32%, and Apo3H 48%). At the high end the mutational load, both Apo3G and Apo3A had 4% of the clones with between 14 and 15 mutations. Sixteen percent of Apo3G mutated clones sustained between 12 and 13 mutations. Apo3H was not able to cause greater than 11 mutations in any clones in this system. The inability of Apo3H to catalyze large numbers of mutations per clone could be due to its preferential bias to deaminate near the 3'-end of DNA (Figure 3.2) rather than the 5'-end of DNA that is single-stranded the longest. Alternatively, different distribution of jumping and sliding events for Apo3H in comparison to Apo3G, could lead to inefficient scanning of the substrate (Figure 3.5 B).

It is unexpected that a non-processive enzyme (Apo3A) could induce so many deaminations in this system. However, the timing of reverse transcription must be slow enough to allow multiple interactions of Apo3A with the (-)DNA in the *lacZα* to induce numerous deaminations (Figure 3.6 B). Yet, this inefficiency is evident in the *prot* region that is single stranded for a shorter time where Apo3A deaminations per clone are less than Apo3G (Figure 3.6 A). Since Apo3H also appears to have a deficiency with respect to the numbers of mutations

per clone, compared to Apo3G (Figure 3.5 and 3.6), it is possible that there is a different distribution of jumping and sliding events (Section 3.2.2) for Apo3H and Apo3G results in their differential ability to induce mutations (Figure 3.5).

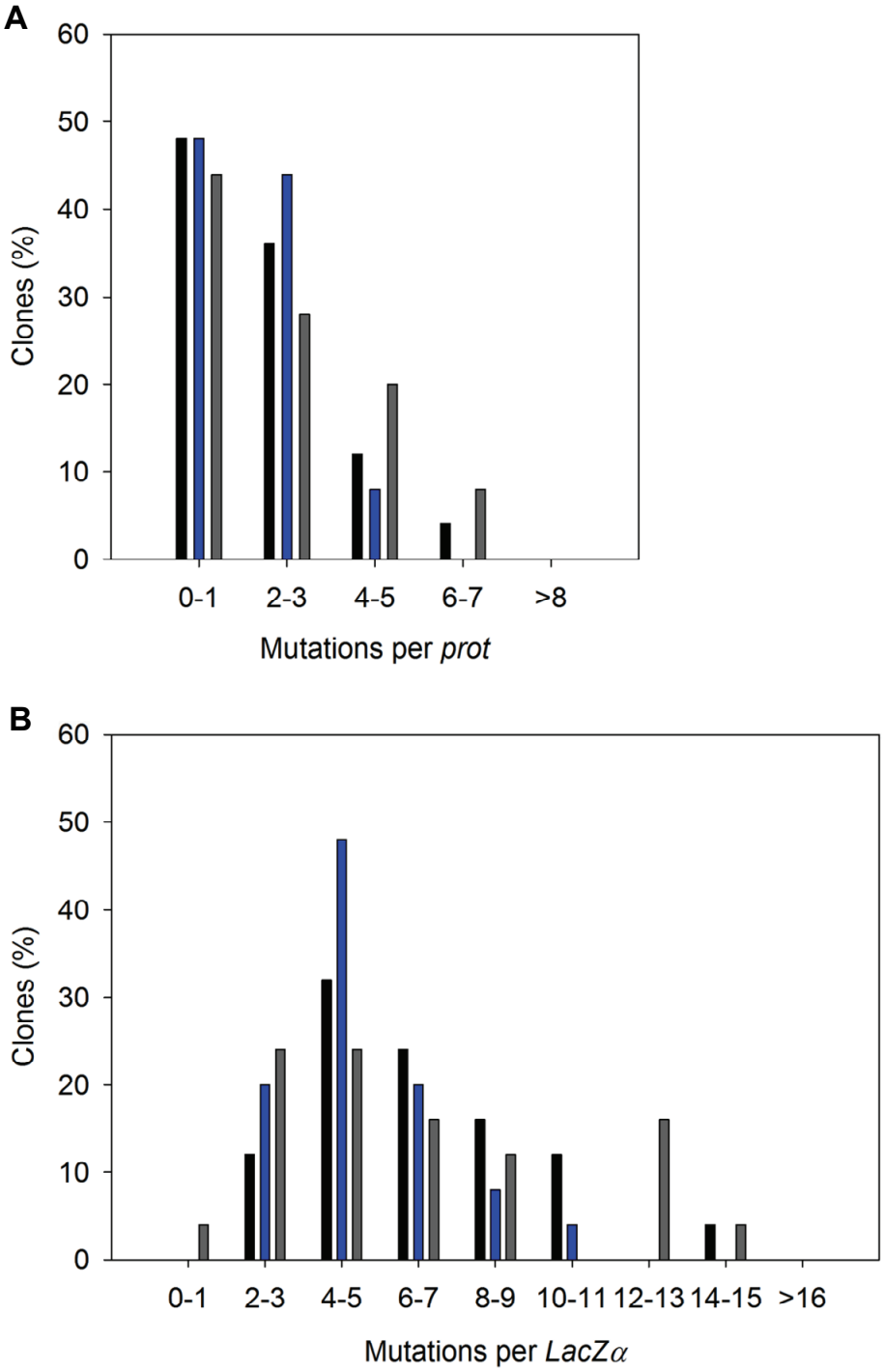


Figure 3.6 **Apo3A-, Apo3H-, and Apo3G-induced mutations per clone.** Plots which illustrate the per clone numbers of G→A mutations that can be obtained in the *prot* (A) or *lacZα* (B) induced by Apo3A (●), Apo3H (●), and Apo3G (●) deaminations. Values were binned for analysis and the bin range is shown on the x-axis. Figure is adapted from Love *et al.* 2012.

3.5.3 Analysis of mutated amino acids and protease activity

To determine the effectiveness of an Apo3 enzyme to functionally inactivate HIV, we analyzed the results of the mutations incurred in the *prot* region of the reverse transcript. The number of mutations sustained by a given clone has been shown as a good indicator of how effective an Apo3 enzyme will be in successfully inactivating an HIV provirus (Harris *et al.*, 2003; Mangeat *et al.*, 2003; Zhang *et al.*, 2003). In a study by Loeb *et al*. each of the amino acids of the protease active site were substituted and tested for activity (Loeb *et al.*, 1989). Loeb *et al*. found that a single amino acid change in the protease active site was sufficient to eliminate its activity. We used the Loeb *et al*. study to compare the resulting amino acid changes in the mutated clones. It was observed that protease genes mutated by Apo3A were predicted to remain active 56% of the time (Table 3.2), both because both the number of mutations was not high enough to cause inactivation and the codon substitutions were silent or did not change protease activity. The specific amino acid changes that Apo3A catalyzed most often were M46I (52%), D35N (28%), and D30N (24%) (Table 3.2), where each of those mutations had no affect on protease activity (Table 3.2). Furthermore, M46I and D30N both conferred the protease with drug resistance as revealed by cross-referencing the mutations with the Stanford HIV Drug Resistance Database. Apo3H did not deaminate any site greater than 36% demonstrating that deaminations catalyzed by this enzyme are more random in nature than the other two Apo3 enzymes tested here (Table 3.2). The top three amino acid substitutions observed for Apo3H were W42STOP (36%), M46I (20%), and D35N (16%) (Table 3.2). Only W42STOP abolished protease activity of the top three amino acids hit by Apo3H. Like Apo3A, Apo3H caused the substitution of M46I which conveys drug resistance. The top four mutations found in clones mutated by Apo3G were G52S (48%), G51R/E (36%), and D25N (32%). Both G52S and G51R/E lead to an inactivation of the protease (Table 3.2). Apo3G also did not cause amino acid substitutions that resulted in protease inhibitor drug resistance with a high efficiency (M46I 4%, and D30N 8%) (Table 3.2).

We summarized the amino acid substitution data by looking at each clone individually to determine the predicted activity of the protease. Here Apo3G was by far the best at inactivating the protease with 72% of the clones coding for an inactive protease, as most of the clones sustained greater number of mutations and those amino acid changes most often eliminated the

activity of the protease (Figure 3.8). Apo3H was able to cause inactivation of the protease in 60% of clones and Apo3A was only able to inactivate 36% of clones sequenced (Figure 3.8). Interestingly, in the protease region, Apo3A was the most efficient at causing a deamination at least once in all of the clones (no mutation 12% for Apo3A, compare 20% for Apo3H and Apo3G) suggesting that non-processive enzymes do interact with more substrate molecules than processive enzymes.

Taken together, these results indicate that the types of mutations that are sustained by a target sequence can be just as important as the number of mutations. Apo3G, which deaminates 5' CCC motifs, had a high level of induced mutagenesis and inactivation (Figure 3.1 B). However, when sparsely mutated clones were examined from Apo3G experiments we found that in the majority of cases, 1 mutation was sufficient for inactivation of protease. This is in contrast to Apo3A and Apo3H which deaminate cytosine in 5'TTC and 5'CTC motifs, respectively, and did not cause high levels of protease inactivation (Table 3.2), despite some clones with high numbers of mutations (Figure 3.6-7). As a result, Apo3A and Apo3H were unable to benefit the "host" of this scenario due to inducing a number of non-inactivating mutations in the protease. Importantly, the deamination motif preference was recapitulated from data on ssDNA (Figure 3.1 A-B and 3.2 A) where the single domain Apo3 enzymes have a greater specific activity on the minimum sequence cytosine motif 5'TC while Apo3G has a preference for 5'CC.

Table 3.2 **Amino acid changes in the HIV-1 protease induced by deaminations by Apo3A, Apo3H, and Apo3G.** Protease enzyme activity was inferred from a complete mutagenesis study performed by Loeb *et al*. (Loeb *et al.*, 1989). Plus (+) indicates that the protease would remain active or partially active given the substitution. Negative (-) indicates that protease activity would be decreased. The Stanford HIV Drug Resistance Database was consulted to determine if a given amino acid substitution would cause resistance. Resistance is indicated by "Yes".

Protease amino acid position	Nucleotide change (underlined)	Amino acid change	Protease enzyme activity	Protease inhibitor resistance	Mutated clones (%)		
					Apo3G	Apo3A	Apo3H
19	CTG→CTA	L→L	+				8
21	GAA→AAA	E→K	+			4	8
25	GAT→AAT	D→N	+		32	4	12
27	GGA→GAA	G→E	-		4	8	12
	GGA→AGA	G→R	-		16		
29	GAT→AAT	D→N	-			8	12
30	GAT→AAT	D→N	+	Yes	8	24	12
34	GAA→AAA	E→K	+		4	20	8
35	GAT→AAT	D→N	+		4	28	16
40	GGG→GGA	G→G	+			4	
	GGG→GAG	G→E	-		8		
	GGG→AGG	G→R	-		8		
	GGG→AAG	G→K	-		8		
42	TGG→TAG	W→STOP	-		16	4	
	TGG→TGA	W→STOP	-			8	36
	TGG→TAA	W→STOP	-				4
46	ATG→ATA	M→I	+	Yes	4	52	20
48	GGG→AGG	G→R	+		16		
	GGG→GGA	G→G	+		4		
49	GGA→AGA	G→R	-		8		
	GGA→GAA	G→E	-				4
	GGA→AAA	G→K	-				4
51	GGA→AAA	G→K	-		4	16	
	GGA→GAA	G→E	-			4	12
	GGA→AGA	G→R	-		32		
52	GGT→AGT	G→S	-		48		

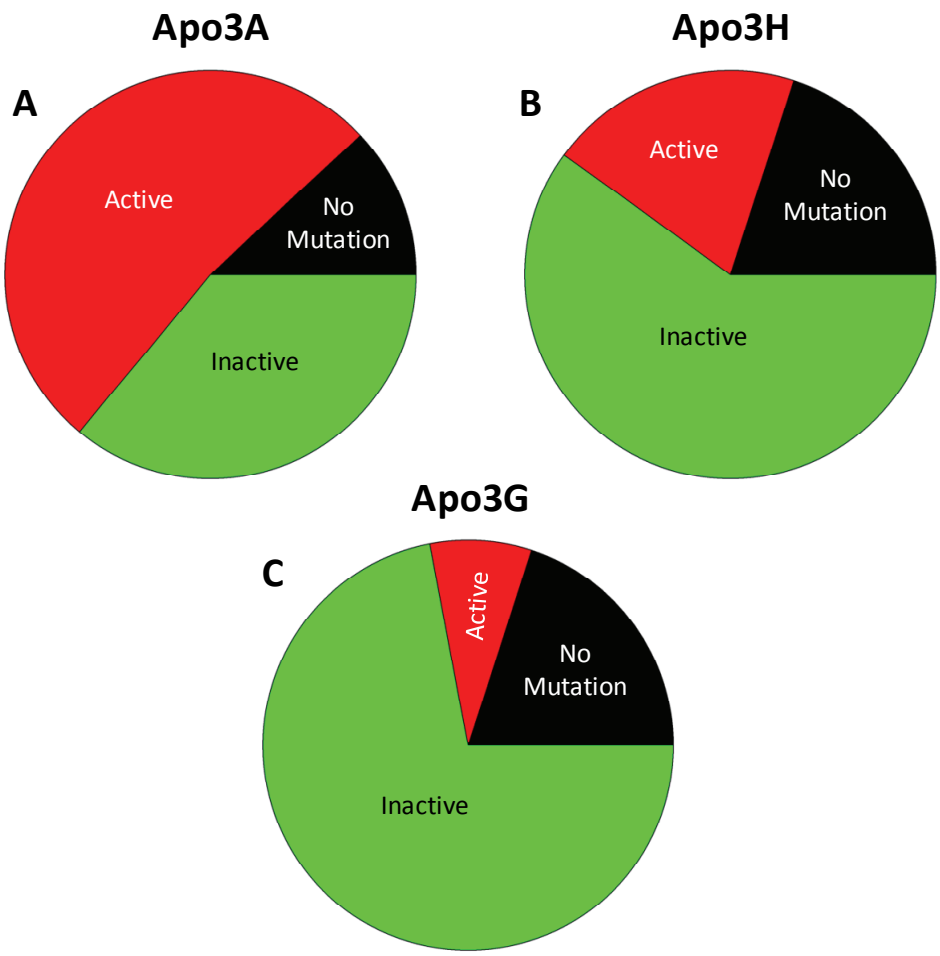


Figure 3.7 **Predicted HIV-1 protease activity effects after Apo3A-, Apo3H-, and Apo3G-induced mutagenesis.** Each *prot* region sequenced was analyzed individually to determine whether deaminations catalyzed by Apo3A (A), Apo3H (B), and Apo3G (C) were able to induce mutations that led to *prot* inactivation (inactive), no inactivation (active), or no mutations. A, Apo3A was able to inactivate the *prot* in 36% of the clones, 12% sustained no mutations, and 52% of clones sustained mutations but remained active. B, Apo3H was able to inactivate the *prot* in 60% of clones, 20% did not have any mutations, and 20% of clones sustained mutations but remained active. C, Apo3G was able to inactivate the *prot* in 72% of the clones, 20% sustained no mutations, and only 8% of the clones sustained mutations and remained active. Figure is adapted from Love *et al.* 2012.

3.6 Deamination during active transcription

Apo3 enzymes can be dangerous for the host cell DNA since they can initiate mutagenesis. The best studied Apo3 enzyme, Apo3G, is found in the cytoplasm in all cell types and is only found in the nucleus when the cell has been irradiated (Nowarski *et al.*, 2012). The single domain Apo3 enzymes, Apo3A and Apo3H, are found in both the nucleus and the cytoplasm (Chen *et al.*, 2006; OhAinle *et al.*, 2008). Nuclear host DNA can become temporarily single stranded when it is being replicated or transcribed. Genes that are transcribed often pose a greater risk of sustaining deaminations by Apo3 type activity as demonstrated by AID (see Section 1.3.4). In particular there is evidence emerging that AID actually interacts with the entirety of the transcribed genome in B cells and would severely damage the host chromosome if it were not for robust DNA repair correcting deaminations in an ongoing process (Hasham *et al.*, 2010).

To test if Apo3A and Apo3H could deaminate during active transcription, we used an experimental system that had been used previously to demonstrate the activity of AID during transcription (Pham *et al.*, 2003). The experiment consisted of designing a double stranded substrate that contained a T7 promoter sequence upstream of a cytosine deamination motif on the non-transcribed strand specific for Apo3A, Apo3H, Apo3G, or AID. The non-transcribed strand is thought to be the substrate for AID within the transcription bubble (Perlot *et al.*, 2008; Pham *et al.*, 2003). Deamination of the cytosine motif can only occur if the T7 RNA polymerase is able to open the double stranded substrate to expose a single stranded deamination motif (data not shown). To detect any nucleotide changes that may have occurred, we sequenced the non-transcribed strand. We set up two modified Sanger sequencing reactions in which either ddA or ddG made up the total A or G in the reaction. This was done to observe the presence of a uracil in the non-transcribed strand directly (ddA) by causing sequence termination at that spot or indirectly by observing sequencing pass through (ddG).

To directly detect the uracil in the DNA substrate, we set up a dideoxynucleotide sequencing reaction where dG, dC, and dT were present and ddA represented the only adenine available. In this way, we would observe a complete stop of all sequenced templates where an A was required for continued synthesis. This would occur if the cytosine were deaminated in 100% of the sequences. In the negative control, we observed a complete stop at the first T of the non-

transcribed strand, which is located downstream from the deamination motif (Figure 3.8 A, *DNA only*). For AID, we observed a stop band in the ddA lane at what was the C position in the DNA sequence, indicating that deamination of the C→U had taken place (Figure 3.8 A, *AID ddA lane*). At this time point not all of the cytosines are converted to uracils (Figure 3.8 B, *30 min*) so we observe sequencing pass through, indicating no deamination (Figure 3.8 A, *downstream T has stop band*). The ratio of integrated gel band intensity between the deaminated and un-deaminated bands was compared to determine the amount of substrate usage. For Apo3A and Apo3H we observed a strong stop band at the location of the cytosine (Figure 3.8 A, *Apo3A and Apo3H*). To accommodate the specific requirements of Apo3A and Apo3H sequence preference, there had to be a T incorporated early in the non-transcribed strand only 1 nt away from the target C (Figure 3.8 A, *Apo3A and Apo3H*). We were still able to detect the difference in size between these extended primers by separating the samples to single nucleotide resolution. Apo3G did not show an easily detectable premature stop band indicating to us that there was little deamination (Figure 3.8 A, *Apo3G*).

To quantify the results of the deamination, we chose to sequence the non-transcribed strand using a ddG reaction. This was necessary because the ddG reaction was able to demonstrate if both of the cytosines in the 5'CTC motif by Apo3H were deaminated (Figure 3.8 A, *Apo3H compare ddA and ddG lanes*). In the Apo3H sequencing reaction, a stop band in the ddG lane would arise when the polymerase was required to incorporate its first cytosine into the extended DNA strand. Here the observed bypass band (Figure 3.8 A, *Apo3H ddG arrow*) is caused by sequencing pass through where the cytosine template has been changed to a U (Figure 3.8 A, *Apo3H ddG*). Apo3H demonstrated 3 stop bands due to its sequence preference (5'CTC); the first represents the un-deaminated templates, the second represents the single deaminated substrates, and the third represents the deaminated substrates where both cytosines were deaminated (Figure 3.8 A, *Apo3H ddG*). Activity recorded for Apo3H was the sum of both of the pass through stop bands. In the ddG lanes for AID and Apo3A, we observed a strong stop band downstream of the deamination motif indicating that the target cytosine had been converted to uracil since sequencing reaction had passed through (Figure 3.8 A, *AID and Apo3A ddG*). For Apo3G, the stop bands observed were only 1 nt apart because sequencing stops at the first C and Apo3G requires a CCC motif (Figure 3.8). Apo3G rarely deaminates more than one C in the CCC motif without an extended incubation time (data not shown).

Apo3A and AID were able to deaminate 50.0% and 48.4% respectively of all cytosines after 30 min (Figure 3.8 B). Apo3G was not able to catalyze deaminations efficiently (Figure 3.8 B, *7.4%, 30 min*) in this system. Apo3H was able to catalyze the deamination of 62.5% of cytosine motifs after 30 min indicating that it was highly active in this system (Figure 3.8 B). We compared the rate of reaction between Apo3H, AID, and Apo3A. Apo3H had the highest rate of reaction at 7% (cytosines deaminated)/min (Figure 3.8 B). AID and Apo3A had rates of 5%/min and 2.5%/min respectively (Figure 3.8 B).

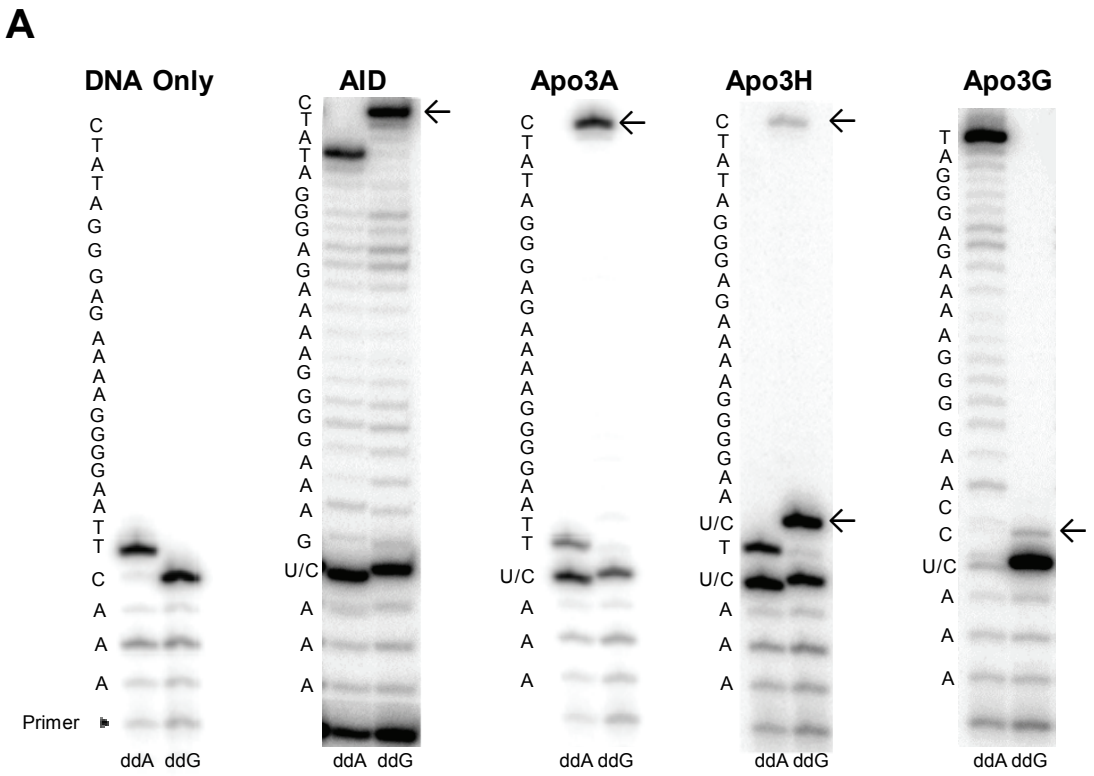

A
DNA Only
AID
Apo3A
Apo3H
Apo3G
Primer
ddA ddG
B
Deamination (%)
Time (min)
Apo3H
AID
Apo3A
Apo3G

Figure 3.8 **Deamination of dsDNA undergoing transcription by Apo3A, Apo3H, AID, and Apo3G.** A, The non-transcribed strand of a dsDNA substrate that underwent T7 RNA polymerase-mediated transcription in the presence of AID, Apo3A, Apo3H, and Apo3G was sequenced using the primer elongation dideoxynucleotide termination assay. The sequence is indicated to the left of each gel. The U/C label indicates the cytosine embedded within the preferred deamination motif for AID, Apo3A, Apo3H and Apo3G. When the C is deaminated to U, a stop band will be present at that position in the ddA lane. In the ddG lane, the presence of a U allows synthesis to bypass the U/C site and the next C in the sequence will have a stop band (arrow) with an intensity that is proportional to the amount of C deamination. DNA only indicates that no deaminase was present and shows the expected stop bands in the absence of deamination. B, time course of deamination of dsDNA during T7 RNA polymerase-mediated transcription. Error bars represent the standard deviation calculated from three independent experiments. Adapted from Love *et al*. 2012.

CHAPTER 4.0

DISCUSSION

4.1 Biochemical determinants for successful restriction of HIV

Restriction of HIV by Apo3 enzymes requires that they interact with the viral genome and induce mutagenesis in a deamination dependent manner (Berger *et al.*, 2011; Yu *et al.*, 2004). In the case of Apo3G, it is packaged into the Δ*vif* HIV virions and upon entering the target cell is able to induce considerable numbers of G to A mutations in the nascent provirus (Harris *et al.*, 2003; Mangeat *et al.*, 2003; Sheehy *et al.*, 2002; Zhang *et al.*, 2003). Apo3A in the target cell can act as a post-entry block and has been demonstrated to cause inhibition of HIV infection in monocytes, macrophages, and myeloid lineage cells (Berger *et al.*, 2011; Koning *et al.*, 2011; Peng *et al.*, 2007). There is however, the potential for the Apo3 enzymes to catalyze sublethal numbers of deaminations that may lead to evolution of the virus (Jern *et al.*, 2009; Kim *et al.*, 2010; Mulder *et al.*, 2008; Pillai *et al.*, 2008; Sadler *et al.*, 2010). In this research, we applied a model HIV replication system that utilized specific nucleic acid structures that mimic HIV replication *in vivo* as well as physiologically relevant concentrations of proteins (Briggs *et al.*, 2004; Coffin, 1997; Xu *et al.*, 2007; Zhu *et al.*, 2003) to assess how well Apo3A, Apo3H and Apo3G can induce deaminations (Figures 2.1 and 3.5-7). We also used synthetic ssDNA substrates to determine the mechanism by which these enzymes scan ssDNA (Figure 3.1 and 3.2). We determined that the characteristic deamination motif preferred by Apo3A, Apo3G, and Apo3H on ssDNA were retained in a dynamic model HIV replication system, but processivity, or lack of, could not be easily discerned (Figure 3.5 and 3.6). Apo3G is highly processive and uses a dual sliding and jumping scanning mode (Chelico *et al.*, 2006). The balance between sliding and jumping was considered crucial for its ability to cause large numbers of mutations (Feng and Chelico, 2011), but it was not known if processivity *per se* was required for inducing mutagenesis during reverse transcription since a non-processive mutant Apo3G has not been developed. We found that while Apo3G was able to induce large numbers of mutations, Apo3A, without being processive, was also able to catalyze many mutations per clone (Figure 3.6). Fewer mutations for Apo3A were found in regions that were single stranded for a small amount of time, suggesting that processivity does play a role, but a seemingly minor role (Figure 3.5). Apo3G was able to generate clones with larger numbers of mutations at a greater frequency than Apo3A, but clones generated by Apo3A were more diverse with less shared mutations than Apo3G (Figure 3.6). Apo3H appeared to be intermediate between Apo3G and Apo3A in terms of the number of mutations per clone and the frequency of highly mutated clones. The data

appear to indicate that this may be because Apo3H uses sliding (85%) more than jumping (15%) (Figure 3.2 and discussion in section 3.2.2). Apo3G uses jumping approximately 60% of the time and slides 40% of the time and this more balanced ratio was found important to have efficient local scanning and reaching many areas along the (-)DNA (Feng and Chelico, 2011).

The most striking result was that no matter how many mutations were induced by Apo3A or Apo3H, their potential for inactivating the HIV protease was much lower than Apo3G due to the propensity of Apo3A and Apo3H to induce missense mutations that change an amino acid within a conserved group (i.e. similar charge or size) as opposed to Apo3G which usually causes missense mutations outside of a conserved amino acid group (Table 3.2). Apo3G has a preference for the 5'CCC motif while Apo3A and Apo3H have a preference for the 5'TC motif and this resulted in a different codon change potential. This resulted in a substantially different amino acid substitution profile in the HIV protease gene. In the sequence that we used, the active site of the *prot* gene of HIV-1, we observed that the sequence motif preference of Apo3G most often results in amino acid changes compared to Apo3A and Apo3H (Table 3.2). In fact, silent mutations were only observed with enzymes that prefer the 5'TC motif (Table 3.2). Clones generated by Apo3A and Apo3H most often coded for an active protease because of the frequency of non-inactivating mutations or silent mutations (Figure 3.7 and Table 3.2).

There appear to be two key differences in the 5'CCC and 5'NTC deamination motifs that account for our observations. The 5'CC motif is found in codons for Gly (Table 3.1, 5'GGN). The deamination would occur on the complementary DNA strand in the context of 5'C<u>C</u>N, 5'CC<u>C</u>, and rarely 5'<u>C</u>CN (underlined C is deaminated) or any combination (Table 3.1). These changes mostly resulted in mutation of the Gly to a charged amino acid (Table 3.1; Arg, Glu, or Lys) which is a non-conservative change that has a high chance to cause gene inactivation, especially if the change takes place near the enzyme active site (Betts and Russel, 2003; Loeb *et al.*, 1989). Further, it has been suggested that Apo3G can result in a high frequency of stop codons being introduced into the HIV coding sequence because the only codon for Trp, 5'TGG is a minimal Apo3G deamination motif in the (-)DNA strand (5'CCA) (Yu *et al.*, 2004). Deamination of the second C would result in the coding sequence being changed to 5'TAG. In our *prot* sequence there is one Trp codon and Apo3G did induce mutagenesis to form a stop codon 16% of the time (Table 3.2). The 5'TC sequence has neither of these features. The 5'TC

sequence most often occurs in the codon for Asp and the deamination-induced mutation changes it to an Asn (Table 3.2). This change from an acidic amino acid (Asp) to its amide (Asn) is a conservative change at position D35 and D25 that maintains the activity of the protease (Betts and Russel, 2003; Loeb *et al.*, 1989). The only other amino acid change Apo3A induced at a high level is the Met to Ile mutation (Table 3.2), which is also considered to be a conservative change (Betts and Russel, 2003; Loeb *et al.*, 1989). Apo3A could cause a stop codon to be formed at position 42 in 4% of clones (Table 3.2), but this is an off-target site, i.e., not 5'TC. However, the Trp codon occurred with a surrounding sequence context of 5'A TGG A, which would read as 5'T CCA T in the (-)DNA, and enabled Apo3A to induce mutagenesis of the codon to 5'TGA in 8% of clones (Table 3.2). The resistance inducing Met to Ile mutation was found in 20% of Apo3H mutated clones (Table 3.2). This suggests that the 5'TC motif is inherently less likely to cause gene inactivation, as demonstrated by our data on Apo3A- and Apo3H-induced mutagenesis of HIV-1 protease (Table 3.2, Figure 3.7). One important note is that Apo3H did induce nonsense mutations in a total of 36% of clones compared to Apo3G with 16% but that did not carry forward to a greater total protease inactivation than Apo3G (Table 3.2, Figure 3.7) . That the deamination motif has a high likelihood to be located in a certain codon allows us to extrapolate our data from the *prot* region (Table 3.2) to infer what would occur during replication of HIV. The data (Table 3.2, Figure 3.7) suggest that Apo3A and Apo3H would be less effective at restricting HIV replication and rather contribute to HIV evolution and drug resistance (Table 3.2, Figure 3.7). It is particularly dangerous that Apo3A is able to induce such a large number of drug resistant mutations at M46I (Table 3.2, 52% of clones) and D30N (Table 3.2, 24% of clones). This is because even if Apo3A were to mutationally inactivate an HIV provirus it would likely harbour these drug resistant mutations and upon integration could act as a repository for viral recombination, as has been demonstrated for Apo3G (Mulder *et al.*, 2008; Smyth *et al.*, 2012). A recent analysis of how Apo3A restricts HIV replication suggests that Apo3A, in addition to or as a result of deaminating (-)DNA, decreases viral transcripts (Berger *et al.*, 2011). This must be an evolutionarily favoured use of Apo3A, in which the provirus formation is blocked and therefore integration is avoided to suppress possible Apo3A-induced viral evolution.

Taken together these results indicate that sequence deamination preference is a higher value determinant of Apo3-mediated HIV-1 inhibition than processivity. It appears that Apo3A and Apo3H might result in HIV evolution rather than its inactivation (Table 3.2 and 3.7). This

may be why Apo3A has a specialized role to inhibit HIV in myeloid cells where it is thought that Apo3A deaminations result in HIV proviral degradation through DNA repair enzyme processing rather than functional inactivation (Berger *et al.*, 2011). If this was not the role for Apo3H it could explain how it was ineffective, leading to its inactivation through evolution.

4.2 The potential of Apo3A and Apo3H to deaminate nuclear DNA

Apo3A and Apo3H can localize to the nucleus where they can come into contact with nuclear DNA (Chen *et al.*, 2006; OhAinle *et al.*, 2008). DNA is only vulnerable to deamination by Apo3 enzymes when it is single stranded, so nuclear DNA is protected when it is not being replicated or transcribed. AID has been demonstrated to cause large amounts of mutagenesis of nuclear DNA in recent reports (Hasham *et al.*, 2010; Liu *et al.*, 2008; Robbiani *et al.*, 2008). AID functions to deaminate nuclear DNA to initiate somatic hypermutation when the antibody genes are being transcribed (Di Noia and Neuberger, 2007; Peled *et al.*, 2008).

Apo3A has been shown to cause cell cycle arrest due to large amounts of nuclear DNA damage, but the mechanism is unknown although it was suggested to be occurring during replication (Landry *et al.*, 2011). Apo3A has also been shown to cause deamination mediated cell death when its interaction partner TRIB3 is silenced, suggesting that human cells exert specific control over Apo3A deamination activity (Aynaud *et al.*, 2012). Here we showed that Apo3A can deaminate DNA during transcription. This could be a mechanism, in addition to accessing ssDNA during nuclear DNA replication by which nuclear DNA is deaminated. Overall it appears there is a genomic cost for a mutationally based retroviral restriction system.

We hypothesized that Apo3H has been evolutionarily destabilized because it was causing deaminations to the host cell or became ineffective at restricting HIV or other retroelements. Although our data suggests Apo3H is not as effective as Apo3G in restriction of HIV (Table 3.2 and Figure 3.5-7), OhAinle *et al*. (2008) suggest that Apo3H has been required for element restriction in the recent past due to evidence of positive selection. The alternative hypothesis appears to be better supported since for Hap I, II and VII, a correlation can be made between the degree of nuclear localization (Hap I > Hap VII > Hap II) and their instability in the cell (Wang *et al.*, 2011; Zhen *et al.*, 2012). Li and Emerman (2011) postulated that a critical residue for stability, Arg 105, mediates an interaction with a cellular protein which could sequester Apo3H

in the cytoplasm, suggesting that nuclear localization is detrimental. In agreement with these hypotheses we tested the potential of transcription as a target for Apo3H mediated deamination and found that indeed Apo3H was highly effective in this regard (Figure 3.2 and Table 3.1). Apo3H was found to catalyze deamination of DNA undergoing transcription with at least a 1.4-fold higher rate a higher rate than Apo3A, Apo3G, and AID (Figure 3.8). The primary target for AID activity is ssDNA exposed during gene transcription (Bachl *et al.*, 2001; Pham *et al.*, 2003). We propose that Apo3H could be too effective in deaminating nuclear DNA that is temporarily single stranded. Future cell-based experiments are necessary to determine the potential detrimental effects of high Apo3H expression (See Section 5.4). But biochemical experiments suggest that Apo3H could be at least as detrimental to cellular DNA as AID. Importantly AID has been suggested to be responsible for 60% of B cell lymphomas (Robbiani *et al.*, 2008).

While the activity for deaminating transcribed DNA was the highest for Apo3H, we did observe a relatively high activity for Apo3A as well (Figure 3.8). We propose two possible explanations for sustained expression of Apo3A in cells coexisting with potential damage. The first explanation is that since the activity of Apo3A is less than AID, DNA repair can easily correct the catalyzed uracils and DNA repair systems do not become saturated, despite the constant presence of Apo3A in the nucleus. The second possible explanation is that a cellular mechanism controls Apo3A activity, similar to AID access to the nucleus being controlled by phosphorylation (Vuong *et al.*, 2009). While this thesis was being written, there was a report published that identified the TRIB3 as able to attenuate Apo3A activity in the nucleus and protects genomic integrity (Aynaud *et al.*, 2012). When the TRIB3 protein was silenced the cells died as a result of hypermutation of their genomes. The TRIB3 protein did not interact with Apo3H (Aynaud *et al.*, 2012), suggesting that cells are missing a control that would suppress the activity of Apo3H. This is in agreement with Li and Emerman (2011) that proposed the stabilizing Arg 105 residue in Apo3H haplotypes mediated an interaction with a cellular protein that re-routed Apo3H localization to the cytoplasm. The Apo3H haplotypes without Arg 105 are primarily nuclear and have a short half life in cells (<30 min) (OhAinle *et al.*, 2008).

Since it was a recent event in evolution that Apo3H lost a nuclear export signal (OhAinle *et al.*, 2008), by acquiring a premature stop codon, perhaps the most rapid way to deal with the acquired genomic instability was to cause the rapid degradation of Apo3H in cells. The

identification of the proposed cellular partner allowing for prolonged Apo3H stability in cells (Li and Emerman, 2011) may shed more light on the evolution of Apo3H in humans and its main function in innate immunity.

4.3 Conclusion

The experiments performed here have demonstrated that there is a large amount of diversity between the Apo3 enzymes and specifically that Apo3G is not a truly representative of the rest of the APOBEC3 family. The Apo3 enzymes were first identified by Apo3G's ability to inhibit HIV and this narrowed the scope of experimentation and discovery surrounding these enzymes. It is now becoming clear that mutagenesis has a role in the immune system both in the adaptive (AID – antibody maturation) and innate (Apo3 – restriction factors) branches. The biochemical data presented here forms a strong platform to test these biochemically-based hypotheses in cell-based experiments.

CHAPTER 5.0

FUTURE STUDIES

5.1 Investigate the relationship between HIV-1 reverse transcriptase fidelity and Apo3

Working with our model reverse transcription system proved to be a rewarding endeavour. We attempted to use this assay to further study processivity by modifying the concentrations and reactions times. The goal was to observe a 15% population mutation frequency (white colonies) by altering the Apo3 enzyme concentration. This would mean that the large majority of clones had arisen from a single encounter with a single Apo3 enzyme. Although this would not represent the *in vivo* situation where Apo3 enzymes could have multiple encounters with viral (-)DNA, we thought it would allow us to further study differences in processivity. From its onset we thought it may be difficult because RT was able to cause around 10% white colonies from inherent replication errors. Once the conditions gave the desired amount of white colonies we sequenced the clones but expected to see mostly random RT-catalyzed errors. The sequences that were returned were unexpected with a greatly reduced number of random RT errors in the presence of the Apo3 enzyme than in its absence. This was confounding, but was repeatable many times and was found to occur in the presence of any of the Apo3 enzymes tested here. It appears that the relationship between the Apo3 enzymes and HIV-1 RT is not one of constant odds, but one where Apo3 enzymes interacting with the (-)DNA may alter the kinetics of RT dNTP incorporation by slowing down the reaction due to steric hindrance (Webb, 2012). This may allow for RT to spend more time in selecting the correct dNTP and result in more accurate replication. If this hypothesis is true then potentially Apo3 enzymes could lead to positive evolution of the HIV-1 virus not just through sublethal numbers of mutations by deamination (Jern *et al.*, 2009; Kim *et al.*, 2010; Mulder *et al.*, 2008; Pillai *et al.*, 2008; Sadler *et al.*, 2010), but by reducing the total number of mutations (including lethal) normally induced by RT.

5.2 Determine level of deamination on the transcribed strand during active transcription

When we performed the active transcription assay we only investigated the level of deamination on the non-transcribed strand. We hypothesized that if any deamination was occurring it would be occurring away from the T7 polymerase, similar to AID (Pham *et al.*, 2003). However, in order for double strand breaks to occur through DNA repair processing, which is the reason why Apo3A- or AID-mediated deaminations are detrimental, some deaminations are expected to occur on the transcribed strand.

5.3 Exploring processivity during active transcription

The active transcription system that we used here was effective in showing differences between Apo3G, a double domain Apo3 enzyme, and the single domain Apo3 enzymes and AID (Figure 3.8). The limitations of the current active transcription assay, compared to the standard deamination assay, are that we did not have a second cytosine motif to investigate any coordinated deamination potential as we demonstrated in the deamination assay (Section 3.1 and 3.2). We did observe more than one deamination with Apo3H due to having two cytosines in the sequence motif preference so it would likely be possible to create a synthetic DNA that possessed two cytosine motifs that could be used for determining if deaminations that occur during gene transcription are processive. If two sequencing primers were used for the same DNA to quantify deamination at each motif separately, the experiment would be quantitative and feasible.

5.4 Determine level of genomic editing induced by Apo3H in tissue culture

The next step to develop our understanding of the level of nuclear DNA damage induced by Apo3H should be investigated in tissue culture. First, we need to demonstrate that Apo3H can deaminate nuclear DNA as our results suggest (Figure 3.8). We plan to use a 293T cell lines that has stable expression of an inhibitor of uracil-N glycosylase (Kaiser and Emerman, 2006).

In the presence or absence of expression of Apo3H (Hap I and VII) we will determine the amount of uracil in nuclear DNA. The detection techniques we plan to use is differential PCR. In this method, different denaturing temperatures are used. If a genome has sustained many G to A mutations then the annealing temperature of the dsDNA is reduced and the PCR reaction can be successful with a lower denaturing temperature. This method has been used to determine the level of Apo3A related editing of nuclear DNA (Suspene *et al.*, 2011a).

CHAPTER 6.0

REFERENCES

Abe, H., Ochi, H., Maekawa, T., Hatakeyama, T., Tsuge, M., Kitamura, S., Kimura, T., Miki, D., Mitsui, F., Hiraga, N.*, et al.* (2009). Effects of structural variations of APOBEC3A and APOBEC3B genes in chronic hepatitis B virus infection. Hepatol Res *39*, 1159-1168.

Aguiar, R.S., Lovsin, N., Tanuri, A., and Peterlin, B.M. (2008). Vpr.A3A chimera inhibits HIV replication. J Biol Chem *283*, 2518-2525.

Albin, J.S., and Harris, R.S. (2010). Interactions of host APOBEC3 restriction factors with HIV-1 in vivo: implications for therapeutics. Expert Rev Mol Med *12*, e4.

Aynaud, M.M., Suspene, R., Vidalain, P.O., Mussil, B., Guetard, D., Tangy, F., Wain-Hobson, S., and Vartanian, J.P. (2012). Human Tribbles 3 protects nuclear DNA from cytidine deamination by APOBEC3A. J Biol Chem *287*, 39182-39192.

Bach, D., Peddi, S., Mangeat, B., Lakkaraju, A., Strub, K., and Trono, D. (2008). Characterization of APOBEC3G binding to 7SL RNA. Retrovirology *5*, 54.

Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M., and Olsson, C. (2001). Increased transcription levels induce higher mutation rates in a hypermutating cell line. J Immunol *166*, 5051-5057.

Beale, R.C., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., and Neuberger, M.S. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J Mol Biol *337*, 585-596.

Bebenek, K., and Kunkel, T.A. (1995). Analyzing fidelity of DNA polymerases. Methods Enzymol *262*, 217-232.

Berg, O.G., Winter, R.B., and von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry *20*, 6929-6948.

Berger, A., Munk, C., Schweizer, M., Cichutek, K., Schule, S., and Flory, E. (2010). Interaction of Vpx and apolipoprotein B mRNA-editing catalytic polypeptide 3 family member A (APOBEC3A) correlates with efficient lentivirus infection of monocytes. J Biol Chem *285*, 12248-12254.

Berger, G., Durand, S., Fargier, G., Nguyen, X.N., Cordeil, S., Bouaziz, S., Muriaux, D., Darlix, J.L., and Cimarelli, A. (2011). APOBEC3A is a specific inhibitor of the early phases of HIV-1 infection in myeloid cells. PLoS Pathog *7*, e1002221.

Betts, M.J., and Russel, R.B. (2003). Amino acid properties and consequences of substitutions (Chichester, UK, John Wiley & Sons).

Biasin, M., Piacentini, L., Lo Caputo, S., Kanari, Y., Magri, G., Trabattoni, D., Naddeo, V., Lopalco, L., Clivio, A., Cesana, E.*, et al.* (2007). Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G: a possible role in the resistance to HIV of HIV-exposed seronegative individuals. J Infect Dis *195*, 960-964.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J., and Malim, M.H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol *14*, 1392-1396.

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006a). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res *34*, 89-95.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006b). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 8780-8785.

Bonvin, M., and Greeve, J. (2007). Effects of point mutations in the cytidine deaminase domains of APOBEC3B on replication and hypermutation of hepatitis B virus in vitro. J Gen Virol *88*, 3270-3274.

Bransteitter, R., Pham, P., Calabrese, P., and Goodman, M.F. (2004). Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J Biol Chem *279*, 51612-51621.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A *100*, 4102-4107.

Briggs, J.A., Simon, M.N., Gross, I., Krausslich, H.G., Fuller, S.D., Vogt, V.M., and Johnson, M.C. (2004). The stoichiometry of Gag protein in HIV-1. Nat Struct Mol Biol *11*, 672-675.

Cagliani, R., Riva, S., Fumagalli, M., Biasin, M., Caputo, S.L., Mazzotta, F., Piacentini, L., Pozzoli, U., Bresolin, N., Clerici, M.*, et al.* (2011). A positively selected APOBEC3H haplotype is associated with natural resistance to HIV-1 infection. Evolution *65*, 3311-3322.

Chelico, L., Pham, P., Calabrese, P., and Goodman, M.F. (2006). APOBEC3G DNA deaminase acts processively 3' --> 5' on single-stranded DNA. Nat Struct Mol Biol *13*, 392-399.

Chelico, L., Pham, P., and Goodman, M.F. (2009). Stochastic properties of processive cytidine DNA deaminases AID and APOBEC3G. Philos Trans R Soc Lond B Biol Sci *364*, 583-593.

Chelico, L., Prochnow, C., Erie, D.A., Chen, X.S., and Goodman, M.F. (2010). Structural model for deoxycytidine deamination mechanisms of the HIV-1 inactivation enzyme APOBEC3G. J Biol Chem *285*, 16195-16205.

Chelico, L., Sacho, E.J., Erie, D.A., and Goodman, M.F. (2008). A model for oligomeric regulation of APOBEC3G cytosine deaminase-dependent restriction of HIV. J Biol Chem *283*, 13780-13791.

Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R., and Weitzman, M.D. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr Biol *16*, 480-485.

Chiu, Y.L., and Greene, W.C. (2008). The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol *26*, 317-353.

Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 15588-15593.

Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997). Retroviruses (Plainview (NY), Cold Spring Harbor Laboratory Press).

Conticello, S.G., Harris, R.S., and Neuberger, M.S. (2003). The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. Curr Biol *13*, 2009-2013.

Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol *22*, 367-377.

Creighton, S., Bloom, L.B., and Goodman, M.F. (1995). Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. Methods Enzymol *262*, 232-256.

Dang, Y., Siew, L.M., Wang, X., Han, Y., Lampen, R., and Zheng, Y.H. (2008). Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem *283*, 11606-11614.

Demorest, Z.L., Li, M., and Harris, R.S. (2011). Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. J Biol Chem *286*, 26568-26575.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

Feng, Y., and Chelico, L. (2011). Intensity of deoxycytidine deamination of HIV-1 proviral DNA by the retroviral restriction factor APOBEC3G is mediated by the noncatalytic domain. J Biol Chem *286*, 11415-11426.

Folta-Stogniew, E., and Williams, K.R. (1999). Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. J Biomol Tech *10*, 51-63.

Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H.*, et al.* (1996). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J Virol *70*, 7013-7029.

Goff, S.P. (2003). Death by deamination: a novel host restriction system for HIV-1. Cell *114*, 281-283.

Goila-Gaur, R., Khan, M.A., Miyagi, E., Kao, S., and Strebel, K. (2007). Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity. Retrovirology *4*, 61.

Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc Natl Acad Sci U S A *100*, 4126-4131.

Hache, G., Liddament, M.T., and Harris, R.S. (2005). The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain. J Biol Chem *280*, 10920-10924.

Halford, S.E., and Marko, J.F. (2004). How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res *32*, 3040-3052.

Harari, A., Ooms, M., Mulder, L.C., and Simon, V. (2009). Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol *83*, 295-303.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803-809.

Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat Immunol *11*, 820-826.

Ho, D.D., and Bieniasz, P.D. (2008). HIV-1 at 25. Cell *133*, 561-565.

Huthoff, H., and Malim, M.H. (2007). Identification of amino acid residues in APOBEC3G required for regulation by human immunodeficiency virus type 1 Vif and Virion encapsidation. J Virol *81*, 3807-3815.

Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J.G. (2006). Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. J Virol *80*, 5992-6002.

Jager, S., Kim, D.Y., Hultquist, J.F., Shindo, K., LaRue, R.S., Kwon, E., Li, M., Anderson, B.D., Yen, L., Stanley, D.*, et al.* (2012). Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. Nature *481*, 371-375.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics *79*, 285-296.

Jern, P., Russell, R.A., Pathak, V.K., and Coffin, J.M. (2009). Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. PLoS Pathog *5*, e1000367.

Jin, X., Brooks, A., Chen, H., Bennett, R., Reichman, R., and Smith, H. (2005). APOBEC3G/CEM15 (hA3G) mRNA levels associate inversely with human immunodeficiency virus viremia. J Virol *79*, 11513-11516.

Kaiser, S.M., and Emerman, M. (2006). Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase Apobec3G. J Virol *80*, 875-882.

Kim, E.Y., Bhattacharya, T., Kunstman, K., Swantek, P., Koning, F.A., Malim, M.H., and Wolinsky, S.M. (2010). Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. J Virol *84*, 10402-10405.

Koning, F.A., Goujon, C., Bauby, H., and Malim, M.H. (2011). Target cell-mediated editing of HIV-1 cDNA by APOBEC3 proteins in human macrophages. J Virol *85*, 13448-13452.

Koning, F.A., Newman, E.N., Kim, E.Y., Kunstman, K.J., Wolinsky, S.M., and Malim, M.H. (2009). Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol *83*, 9474-9485.

Land, A.M., Ball, T.B., Luo, M., Pilon, R., Sandstrom, P., Embree, J.E., Wachihi, C., Kimani, J., and Plummer, F.A. (2008). Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. J Virol *82*, 8172-8182.

Landry, S., Narvaiza, I., Linfesty, D.C., and Weitzman, M.D. (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. EMBO Rep *12*, 444-450.

Larijani, M., and Martin, A. (2012). The biochemistry of activation-induced deaminase and its physiological functions. Semin Immunol.

LaRue, R.S., Andresdottir, V., Blanchard, Y., Conticello, S.G., Derse, D., Emerman, M., Greene, W.C., Jonsson, S.R., Landau, N.R., Lochelt, M.*, et al.* (2009). Guidelines for naming nonprimate APOBEC3 genes and proteins. J Virol *83*, 494-497.

Li, M.M., and Emerman, M. (2011). Polymorphism in human APOBEC3H affects a phenotype dominant for subcellular localization and antiviral activity. J Virol *85*, 8197-8207.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841-845.

Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E., and Hutchison, C.A., 3rd (1989). Complete mutagenesis of the HIV-1 protease. Nature *340*, 397-400.

Love, R.P., Xu, H., and Chelico, L. (2012). Biochemical Analysis of Hypermutation by the Deoxycytidine Deaminase APOBEC3A. J Biol Chem *287*, 30812-30822.

Madsen, P., Anant, S., Rasmussen, H.H., Gromov, P., Vorum, H., Dumanski, J.P., Tommerup, N., Collins, J.E., Wright, C.L., Dunham, I.*, et al.* (1999). Psoriasis upregulated phorbolin-1 shares structural but not functional similarity to the mRNA-editing protein apobec-1. J Invest Dermatol *113*, 162-169.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99-103.

Marin, M., Golem, S., Rose, K.M., Kozak, S.L., and Kabat, D. (2008). Human immunodeficiency virus type 1 Vif functionally interacts with diverse APOBEC3 cytidine

deaminases and moves with them between cytoplasmic sites of mRNA metabolism. J Virol *82*, 987-998.

Marin, M., Rose, K.M., Kozak, S.L., and Kabat, D. (2003). HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med *9*, 1398-1403.

McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. Proc Natl Acad Sci U S A *103*, 8798-8803.

McIlwraith, M.J., Vaisman, A., Liu, Y., Fanning, E., Woodgate, R., and West, S.C. (2005). Human DNA polymerase eta promotes DNA synthesis from strand invasion intermediates of homologous recombination. Mol Cell *20*, 783-792.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G., and Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem *281*, 22161-22172.

Mulder, L.C., Harari, A., and Simon, V. (2008). Cytidine deamination induced HIV-1 drug resistance. Proc Natl Acad Sci U S A *105*, 5501-5506.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem *274*, 18470-18476.

Muschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., and Kuppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J Exp Med *192*, 1833-1840.

Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K., and Landau, N.R. (2005). Complementary function of the two catalytic domains of APOBEC3G. Virology *333*, 374-386.

Nowarski, R., Wilner, O.I., Cheshin, O., Shahar, O.D., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R.S., Goldberg, M.*, et al.* (2012). APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. Blood.

OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S., and Emerman, M. (2008). Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe *4*, 249-259.

OhAinle, M., Kerns, J.A., Malik, H.S., and Emerman, M. (2006). Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol *80*, 3853-3862.

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J Exp Med *197*, 1173-1181.

Pace, C., Keller, J., Nolan, D., James, I., Gaudieri, S., Moore, C., and Mallal, S. (2006). Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. J Virol *80*, 9259-9269.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341-346.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. Annu Rev Immunol *26*, 481-511.

Peng, G., Greenwell-Wild, T., Nares, S., Jin, W., Lei, K.J., Rangel, Z.G., Munson, P.J., and Wahl, S.M. (2007). Myeloid differentiation and susceptibility to HIV-1 are linked to APOBEC3 expression. Blood *110*, 393-400.

Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. Proc Natl Acad Sci U S A *105*, 3843-3848.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Pham, P., Smolka, M.B., Calabrese, P., Landolph, A., Zhang, K., Zhou, H., and Goodman, M.F. (2008). Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. J Biol Chem *283*, 17428-17439.

Pillai, S.K., Wong, J.K., and Barbour, J.D. (2008). Turning up the volume on mutational pressure: is more of a good thing always better? (A case study of HIV-1 Vif and APOBEC3). Retrovirology *5*, 26.

Pomerantz, R.J. (2002). HIV: a tough viral nut to crack. Nature *418*, 594-595.

Refsland, E.W., Stenglein, M.D., Shindo, K., Albin, J.S., Brown, W.L., and Harris, R.S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res *38*, 4274-4284.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A.*, et al.* (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565-575.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al.* (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R.*, et al.* (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell *36*, 631-641.

Sadler, H.A., Stenglein, M.D., Harris, R.S., and Mansky, L.M. (2010). APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. J Virol *84*, 7396-7404.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol *2*, E275.

Senavirathne, G., Jaszczur, M., Auerbach, P.A., Upton, T.G., Chelico, L., Goodman, M.F., and Rueda, D. (2012). Single-stranded DNA Scanning and Deamination by APOBEC3G Cytidine Deaminase at Single Molecule Resolution. J Biol Chem *287*, 15826-15835.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646-650.

Sheehy, A.M., Gaddis, N.C., and Malim, M.H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med *9*, 1404-1407.

Smyth, R.P., Davenport, M.P., and Mak, J. (2012). The origin of genetic diversity in HIV-1. Virus Res *169*, 415-429.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu Rev Immunol *26*, 261-292.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol *17*, 222-229.

Stephens, C.R., and Waelbroeck, H. (1999). Codon bias and mutability in HIV sequences. J Mol Evol *48*, 390-397.

Stopak, K., de Noronha, C., Yonemoto, W., and Greene, W.C. (2003). HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell *12*, 591-601.

Suspene, R., Aynaud, M.M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J.P., and Wain-Hobson, S. (2011a). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. Proc Natl Acad Sci U S A *108*, 4858-4863.

Suspene, R., Aynaud, M.M., Koch, S., Pasdeloup, D., Labetoulle, M., Gaertner, B., Vartanian, J.P., Meyerhans, A., and Wain-Hobson, S. (2011b). Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. J Virol *85*, 7594-7602.

Suspene, R., Rusniok, C., Vartanian, J.P., and Wain-Hobson, S. (2006). Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucleic Acids Res *34*, 4677-4684.

Suspene, R., Sommer, P., Henry, M., Ferris, S., Guetard, D., Pochet, S., Chester, A., Navaratnam, N., Wain-Hobson, S., and Vartanian, J.P. (2004). APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. Nucleic Acids Res *32*, 2421-2429.

Tan, L., Sarkis, P.T., Wang, T., Tian, C., and Yu, X.F. (2009). Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. FASEB J *23*, 279-287.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621-629.

Ulenga, N.K., Sarr, A.D., Thakore-Meloni, S., Sankale, J.L., Eisen, G., and Kanki, P.J. (2008). Relationship between human immunodeficiency type 1 infection and expression of human APOBEC3G and APOBEC3F. J Infect Dis *198*, 486-492.

Vartanian, J.P., Guetard, D., Henry, M., and Wain-Hobson, S. (2008). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science *320*, 230-233.

Vazquez-Perez, J.A., Ormsby, C.E., Hernandez-Juan, R., Torres, K.J., and Reyes-Teran, G. (2009). APOBEC3G mRNA expression in exposed seronegative and early stage HIV infected individuals decreases with removal of exposure and with disease progression. Retrovirology *6*, 23.

Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. Nat Immunol *10*, 420-426.

Wang, T., Tian, C., Zhang, W., Luo, K., Sarkis, P.T., Yu, L., Liu, B., Yu, Y., and Yu, X.F. (2007). 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. J Virol *81*, 13112-13124.

Wang, W., and Malcolm, B.A. (1999). Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. Biotechniques *26*, 680-682.

Wang, X., Abudu, A., Son, S., Dang, Y., Venta, P.J., and Zheng, Y.H. (2011). Analysis of Human APOBEC3H Haplotypes and Anti-Human Immunodeficiency Virus Type-1 Activity. J Virol.

Webb, J. (2012). The APOBEC3G Deamination Independent Mode of HIV Inhibition. In Microbiology and Immunology (Saskatoon SK, University of Saskatchewan).

Wedekind, J.E., Dance, G.S., Sowden, M.P., and Smith, H.C. (2003). Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet *19*, 207-216.

Wyatt, P.J. (1993). Light-Scattering and the Absolute Characterization of Macromolecules. Analytica Chimica Acta *272*, 1-40.

Xu, H., Chertova, E., Chen, J., Ott, D.E., Roser, J.D., Hu, W.S., and Pathak, V.K. (2007). Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions. Virology *360*, 247-256.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat Immunol *12*, 62-69.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol *11*, 435-442.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature *424*, 94-98.

Zhang, W., Du, J., Evans, S.L., Yu, Y., and Yu, X.F. (2012). T-cell differentiation factor CBF-beta regulates HIV-1 Vif-mediated evasion of host restriction. Nature *481*, 376-379.

Zhen, A., Du, J., Zhou, X., Xiong, Y., and Yu, X.F. (2012). Reduced APOBEC3H variant anti-viral activities are associated with altered RNA binding activities. PLoS One *7*, e38771.

Zhu, P., Chertova, E., Bess, J., Jr., Lifson, J.D., Arthur, L.O., Liu, J., Taylor, K.A., and Roux, K.H. (2003). Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proc Natl Acad Sci U S A *100*, 15812-15817.

CHAPTER 7.0

APPENDIX

Appendix Item 7.1 **Mutation possibilities as revealed in context of HIV reverse transcription**. Deaminations occur on the non-coding strand. Codons highlighted in red indicate no change to amino acid, green indicates change in conservative amino acid group, white indicates same conservative amino acid group (Loeb *et al.*, 1989). Star indicated preferred codon usage pattern for HIV (Stephens and Waelbroeck, 1999).

5'TC motif within codon

Aspartic Acid	*GAU	→	AAU	Asparagine
Aspartic Acid	GAC	→	AAC	Asparagine
Glutamic Acid	*GAA	→	AAA	Lysine
Glutamic Acid	GAG	→	AAG	Lysine
Stop (Opal)	UGA	→	UAA	Stop (Ochre)
Arginine	CGA	→	CAA	Glutamine
Arginine	*AGA	→	AAA	Lysine
Glycine	*GGA	→	GAA	Glutamic Acid

5'CC motif within codon

Tryptophan	*UGG	→	UAG	Stop (Amber)
Arginine	CGG	→	CAG	Glutamine
Arginine	AGG	→	AAG	Lysine
Glycine	GGU	→	AGU	Serine
Glycine	GGC	→	AGC	Serine
Glycine	*GGA	→	AGA	Arginine
Glycine	GGG	→	AGG	Arginine

5'(T/C) preceding codon

Leucine	UUG(A/G)	→	UUA	Leucine
Leucine	CUG(A/G)	→	CUA	Leucine
Methionine	AUG(A/G)	→	AUA	Isoleucine
Valine	GUG(A/G)	→	GUA	Valine
Serine	UCG(A/G)	→	UCA	Serine
Proline	CCG(A/G)	→	CCA	Proline
Threonine	ACG(A/G)	→	ACA	Threonine
Alanine	GCG(A/G)	→	GCA	Alanine
Stop (Amber)	UAG(A/G)	→	UAA	Stop (Ochre)
Glutamine	CAG(A/G)	→	CAA	Glutamine
Lysine	AAG(A/G)	→	AAA	Lysine
Glutamic acid	GAG(A/G)	→	GAA	Glutamic acid
Tryptophan	UGG(A/G)	→	UGA	Stop (Opal)
Arginine	CGG(A/G)	→	CGA	Arginine
Arginine	AGG(A/G)	→	AGA	Arginine
Glycine	GGG(A/G)	→	GGA	Glycine

No possible deamination

Phenylalanine	UUU, UUC
Isoleucine	AUU, AUC
Proline	CCU, CCC, CCA
Tyrosine	UAU, UAC
Stop (Ochre)	UAA
Histidine	CAU, CAC
Asparagine	AAU, AAC

Appendix Item 7.2 **Calculations used in determination of processivity factor**

DEFINING VARIABLES

5' deamination

(5') = (5' deamination band intensity)/(total band intensities)

3' deamination

(3') = (3' deamination band intensity)/(total band intensities)

Actual correlated deamination

(5'&3') = (double deamination band intensity)/(total band intensities)

Total 5' deamination

$(5')_C$ = (5' & 3') + (5')

Total 3' deamination

$(3')_C$ = (5' & 3') + (3')

Expected Value of correlated deamination = $(5')_C$ x $(3')_C$

Processivity Factor = Actual/Expected = (5' & 3')/[$(5')_C$ x $(3')_C$]